

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 14, 2003, 16:16:55 ; Search time 86 Seconds  
(without alignments)  
1225.517 Million cell updates/sec

Title: US-09-590-991-6

Perfect score: 4398  
Sequence: 1 YPVVLADTSSSEDALNISDK.....PGNNKSTLVINGKSTYGY 664

Scoring table: BLOSUM30  
Gapop 1.0 , Gapext 1.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: A\_GeneSeq\_19Jun03.\*  
2: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
3: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
4: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
5: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
6: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
7: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
8: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*  
24: /SIDSL/gcgdata/geneSeq/geneSeq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4398	100.0	664	22	AA848342
2	4398	100.0	2120	21	AA81710
3	4393	99.9	2140	24	ABU01020
4	4171	94.8	1007	19	AA61246
5	4171	94.8	1007	23	ABP54664
6	1549	35.2	2835	23	ABP8574
7	1541	35.0	1475	11	AA808221
8	1525	34.7	2710	17	AA85016
9	1525	34.7	2710	19	AA868387

10	1511	34.4	1530	23	AB847329	Listeria monocytog
11	1509	34.3	1938	17	AAW09255	Bacillus alkaline
12	1509	34.3	1938	20	AAW73553	Full length Pullul
13	1506	34.2	2274	23	ABP30377	Streptococcus poly
14	1506	34.2	2278	23	ABP28340	Streptococcus poly
15	1502	34.2	2057	21	AA10667	L. mesenteroides a
16	1496	34.0	1450	23	AB853597	Lactococcus lactis
17	1493	33.9	1968	12	AA10559	Mutant protease (A
18	1490	33.9	1968	12	AA10561	Mutant protease (d
19	1489	33.9	1974	12	AA10940	Mutant protease (d
20	1488	33.8	3097	22	AB862967	Drosophila melanog
21	1486	33.8	1482	22	ABG07451	Novel human diagno
22	1485	33.8	7107	22	AB858144	Drosophila melanog
23	1484	33.7	1560	10	AA894145	S. cremoris proteol
24	1483	33.7	1962	12	AA10557	Mutant protease (A
25	1482	33.7	1962	12	AA10558	Mutant protease (A
26	1481	33.7	1941	23	ABP65509	Bifidobacterium 10
27	1480	33.7	1781	23	AAU74519	Lactobacillus reut
28	1479	33.6	1962	12	AA10560	Mutant protease (K
29	1479	33.6	1962	12	AA10561	Mutant protease (N
30	1479	33.6	1962	12	AA10563	Mutant protease (K
31	1479	33.6	2032	20	AAU00238	Enterococcus faeca
32	1479	33.6	2032	20	AAU00240	Enterococcus faeca
33	1479	33.6	2032	20	AAU00242	Enterococcus faeca
34	1479	33.6	2032	23	ABP43457	E faecalis EF123 p
35	1479	33.6	2032	23	ABP43459	E faecalis EF125 p
36	1479	33.6	2032	23	ABP43461	Enterococcus faeca
37	1479	33.6	2032	24	ABU13736	Enterococcus faeca
38	1479	33.6	2032	24	ABU13738	Enterococcus faeca
39	1479	33.6	2032	24	ABU13740	Enterococcus faeca
40	1478	33.6	1959	12	AA10562	Mutant protease (d
41	1478	33.6	2366	17	AA85011	C. difficile toxin
42	1478	33.6	2366	19	AA868388	Clostridium diffic
43	1477	33.6	1959	23	AB849411	Listeria monocytog
44	1476	33.6	1584	20	AA133727	Photobacterium lum
45	1476	33.6	2681	24	ABU19025	Pathogen specific

#### ALIGNMENTS

RESULT 1  
AAB48342  
AAB48342 standard; Protein; 664 AA.  
XX  
AC AAB48342;  
XX  
DT 20-APR-2001 (first entry)  
XX  
DE S. pneumoniae Sp128 polypeptide.  
XX  
KW Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;  
KW bronchial; lung; blood; infection; immune response; immunotherapy;  
KW antibacterial; auditory; vaccine.  
XX  
OS Streptococcus pneumoniae.  
XX  
PN W0200076540-A2.  
XX  
PD 21-DEC-2000.  
XX  
PF 09-JUN-2000; 2000MO-US15925.  
XX  
PR 10-JUN-1999; 99US-0138453.  
XX  
PA (MED1-) MED IMMUNE INC.  
XX  
PI Adamou JE, Choi GH;  
XX  
DR WPI, 2001-112197/12.  
XX  
DR N-PSDB; AAC84741.  
XX  
PT New vaccines comprising Sp128 or Sp130 polypeptides, for treating and

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 PT blood infections -

PS Claim 4; Page 47-50; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Spi28 and Spi30  
 CC from *S. pneumoniae*. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by Streptococcus, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the *S. pneumoniae* Spi28 polypeptide.

XX Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;

Best Local Similarity 100.0%; Pred. No. 2.3e-77; Mismatches 0; Indels 0; Gaps 0;

Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNIDKEVAENKEKHENIHSAHETSODPFKEKTAIVIKEKEVSKN 60  
 DB 1 YPVVLADTSSSEDALNIDKEVAENKEKHENIHSAHETSODPFKEKTAIVIKEKEVSKN 60  
 QY 61 PVIDNNTSNEAKIKEENSKNSQGDYTDSPVNKNTENPKKEDKVVYIAEFKDGESGEKAI 120  
 DB 61 PVIDNNTSNEAKIKEENSKNSQGDYTDSPVNKNTENPKKEDKVVYIAEFKDGESGEKAI 120  
 QY 121 KELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOIGISISVERAKOVPMNNHARKEI 180  
 DB 121 KELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOIGISISVERAKOVPMNNHARKEI 180  
 QY 181 GVEBAIDVLSKINAFPGKMFDRGMVINSIDTGTDRYRHKARIDDDAASRPFKEDLKG 240  
 DB 181 GVEBAIDVLSKINAFPGKMFDRGMVINSIDTGTDRYRHKARIDDDAASRPFKEDLKG 240  
 QY 241 TDKNYMSLDKIPHANVYNGGKITVEKTDGGRDYDPHGMIAGILANDTBODIKNNG 300  
 DB 241 TDKNYMSLDKIPHANVYNGGKITVEKTDGGRDYDPHGMIAGILANDTBODIKNNG 300  
 QY 301 IDGAPNAQIFSYKMYSDAGSGFADDETFPHAIEDSIKANDVVSSSGFTGTGVGEKY 360  
 DB 301 IDGAPNAQIFSYKMYSDAGSGFADDETFPHAIEDSIKANDVVSSSGFTGTGVGEKY 360  
 QY 361 WQAIKALKRAGIPMVVATGNTATSSASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAIAS 420  
 DB 361 WQAIKALKRAGIPMVVATGNTATSSASSSSWDLVANNHLMKMTDGTGNVTRTAHEDAIAS 420  
 QY 421 AKNQVBERDKVNIIGESFRYRNIGAFPPKSKITTNEDGTAPSKLKFYIIGKQDQDLIG 480  
 DB 421 AKNQVBERDKVNIIGESFRYRNIGAFPPKSKITTNEDGTAPSKLKFYIIGKQDQDLIG 480  
 QY 481 LDLGKIAVMRIYTKDKLNAFKKAMDKGARIMVNVVYNNRDNMTLPMAGYEADBG 540  
 DB 481 LDLGKIAVMRIYTKDKLNAFKKAMDKGARIMVNVVYNNRDNMTLPMAGYEADBG 540  
 QY 541 TKSQVFSISGDDGVGLAMNINPDKTEVYRNNKEDFKDLRQYPIIDMESFNSNKNPNVD 600  
 DB 541 TKSQVFSISGDDGVGLAMNINPDKTEVYRNNKEDFKDLRQYPIIDMESFNSNKNPNVD 600  
 QY 601 EKEIDFKAPPTDKELKYEDIIIVPAGSTSWGPRIDLKLPDVSAGKNIKSTLVYINGKS 660  
 DB 601 EKEIDFKAPPTDKELKYEDIIIVPAGSTSWGPRIDLKLPDVSAGKNIKSTLVYINGKS 660  
 QY 661 TYGY 664  
 DB 661 TYGY 664

RESULT 2  
 ID AAY81710 standard; Protein: 2120 AA.

XX AAY81710;

XX 02-JUN-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID3.

KM Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KM bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KM kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KM pneumococcal septicemia; sinusitis; meningitis; therapy.

OS Streptococcus pneumoniae.

XX WO200006738-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02452.

XX 27-JUL-1998; 98GB-0016336.

XX 19-MAR-1999; 99US-0125329.

PA (MICR-) MICROBIAL TECHNICS LTD.

PI Le Page RWF, Wells JM, Hamifly SB, Hanebro PM;

XX WPI; 2000-195301/17.

DR N-PSDB; AAZ91806.

XX Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections -

XX Claim 2; Page 41-42; 76pp; English.

XX This sequence represents a Streptococcus pneumoniae protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful for the detection or  
 CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*  
 CC infection. As the sequences can be used to treat *S. pneumoniae* infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
 CC meningitis.

XX Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;

Best Local Similarity 100.0%; Pred. No. 4.5e-75; Mismatches 0; Indels 0; Gaps 0;

Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNIDKEVAENKEKHENIHSAHETSODPFKEKTAIVIKEKEVSKN 60  
 DB 1 YPVVLADTSSSEDALNIDKEVAENKEKHENIHSAHETSODPFKEKTAIVIKEKEVSKN 60  
 QY 61 PVIDNNTSNEAKIKEENSKNSQGDYTDSPVNKNTENPKKEDKVVYIAEFKDGESGEKAI 120  
 DB 61 PVIDNNTSNEAKIKEENSKNSQGDYTDSPVNKNTENPKKEDKVVYIAEFKDGESGEKAI 120  
 QY 121 KELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOIGISISVERAKOVPMNNHARKEI 180  
 DB 121 KELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOIGISISVERAKOVPMNNHARKEI 180

```

QY 181 GVEBAIDYLSINAPFGKNPDGRGMVINSIDTGTDRHKKARIDDDAKASMRFKEDLKG 240
DB 181 GVEBAIDYLSINAPFGKNPDGRGMVINSIDTGTDRHKKARIDDDAKASMRFKEDLKG 240
QY 241 TDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNTEODIKNPFG 300
DB 241 TDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNTEODIKNPFG 300
QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVYVSSGFTGTLVGEKY 360
DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVYVSSGFTGTLVGEKY 360
QY 361 WQAIRALRKAGIPMVVATGNVATSSASSSWDLVANNHUKMTDGNVTRTAHEDAIAVAS 420
DB 361 WQAIRALRKAGIPMVVATGNVATSSASSSWDLVANNHUKMTDGNVTRTAHEDAIAVAS 420
QY 421 AKQNTVEFDKNIGESFKYRNIGAFPDKSKITTNEGDTAPSKLKVYIGKGDODLIG 480
DB 421 AKQNTVEFDKNIGESFKYRNIGAFPDKSKITTNEGDTAPSKLKVYIGKGDODLIG 480
QY 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGAIAIMVAVTVYNNRDNNTLPRMGYEADG 540
DB 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGAIAIMVAVTVYNNRDNNTLPRMGYEADG 540
QY 541 TKSQVFSISGDDGYKLMNMINDPKTEYKRNKEDFPDKLEQYYPIDMESFNKPNVGD 600
DB 541 TKSQVFSISGDDGYKLMNMINDPKTEYKRNKEDFPDKLEQYYPIDMESFNKPNVGD 600
QY 601 EKEIDFPAPDTELYKEDIIYVAGSTSWGPRIDLKLPVSAFGNISTLNVINGKS 660
DB 601 EKEIDFPAPDTELYKEDIIYVAGSTSWGPRIDLKLPVSAFGNISTLNVINGKS 660
QY 661 TYGY 664
DB 661 TYGY 664

```

RESULT 3  
ABU01020  
ID ABU01020 standard; Protein: 2140 AA.

```

XX AC ABU01020;
XX DT 11-FEB-2003 (first entry)
XX DE S. pneumoniae type 4 strain protein from coding region #590.
XX KM Bacterial meningitis; pneumonia; sepsis; otitis media;
XX KM ear infection; antiinflammatory; antibacterial; immunostimulant;
XX KM auditory; respiratory; gene therapy; vaccine.
XX OS Streptococcus pneumoniae type 4 strain.
XX PN W0200277021-A2.
XX PD 03-OCT-2002.
XX PF 27-MAR-2002; 2002W0-IB02163.
XX PR 27-MAR-2001; 2001GB-0007658.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Malignant V, Tectelin H, Fraser C;
XX DR WPI; 2003-040579/03.
XX DR N-PSDB; ABX06302.
XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX PT useful as medicaments for treating or preventing a disease or infection
XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
XX PT or ear infection

```

```

XX XX
XX Claim 1; SEQ ID No 1180; 56pp; English.
PS
XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,
XX expressed from 2469 of 2489 identified DNA coding regions from the
XX Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX AB556454. Also included are an antibody which binds one of the
XX proteins, treating a patient by administering the protein, DNA or
XX antibody (in a composition), a kit comprising first and second primers,
XX which are the nucleic acid cited above or fragments between nucleotides
XX 8-100 of a sequence not defined in the specification, for amplifying a
XX target sequence contained within a Streptococcus nucleic acid sequence,
XX where the first primer is substantially complementary to the target
XX sequence and the second primer is substantially complementary to the
XX complement of the target sequence, and where the pairs of the primers
XX having substantial complementarity define the termini of the target
XX sequence to be amplified, assay comprising contacting a test compound
XX with the protein, and determining whether the test compound binds to the
XX protein and a Streptococcus pneumoniae bacterium, where one or more
XX genes encoding the proteins has been rendered inactive. The proteins,
XX nucleic acid molecules, antibody and compositions are useful as
XX medicaments for treating or preventing a disease or infection due to
XX streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX sepsis, otitis media or ear infection. They are also useful in developing
XX vaccines, diagnostics and antibiotics. The methods are useful for
XX identifying immunodominant proteins. The present sequence is one of
XX the 2469 proteins expressed by the identified coding regions from the
XX genomic sequence.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2140 AA;
SQ

```

Query Match 99.9%; Score 4393; DB 24; Length 2140;  
Best Local Similarity 99.6%; Pred. No. 66-75;  
Matches 664; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

```

QY 1 Y-P--VVLADTSSSEDALNISDKKVAENKKEHNIHSAHETSQDFEKKTAIVKEKEV 57
DB 19 YAPNEVVLADTSSSEDALNISDKKVAENKKEHNIHSAHETSQDFEKKTAIVKEKEV 78
QY 58 SKRPVINDNTSNEAKIKENSNSKQDYYDSFVNKNTENPKEDKRVYIAEFDKESGE 117
DB 79 SKRPVINDNTSNEAKIKENSNSKQDYYDSFVNKNTENPKEDKRVYIAEFDKESGE 138
QY 118 KAIKELSSLKNTKVLTYDRIFNGSAIETTPDMLDKIQLEGISVERAOKVOPMNNHAR 177
DB 139 KAIKELSSLKNTKVLTYDRIFNGSAIETTPDMLDKIQLEGISVERAOKVOPMNNHAR 198
QY 178 KEIGVEBAIDYLSINAPFGKNPDGRGMVINSIDTGTDRHKKARIDDDAKASMRFKED 237
DB 199 KEIGVEBAIDYLSINAPFGKNPDGRGMVINSIDTGTDRHKKARIDDDAKASMRFKED 258
QY 238 LKGTDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNTEODIKN 297
DB 259 LKGTDKRYWLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNTEODIKN 318
QY 298 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVYVSSGFTGTLVGE 357
DB 319 FNGIDGIAPNAQIFSYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVYVSSGFTGTLVGE 378
QY 358 EKYWQAIRALRKAGIPMVVATGNVATSSASSSWDLVANNHUKMTDGNVTRTAHEDAIA 417
DB 379 EKYWQAIRALRKAGIPMVVATGNVATSSASSSWDLVANNHUKMTDGNVTRTAHEDAIA 438
QY 418 VASAKQNTVEFDKNIGESFKYRNIGAFPDKSKITTNEGDTAPSKLKVYIGKGDOD 477
DB 439 VASAKQNTVEFDKNIGESFKYRNIGAFPDKSKITTNEGDTAPSKLKVYIGKGDOD 498

```

```

QY 478 LIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNTLPAWGYA 537
   |||||
Db 499 LIGDLRKRKIAVMDRIYTKDLKNAFKKAMDKGARAIMVNTVNYNRDNTLPAWGYA 558
   |||||
QY 538 DEGTYSQVFSISGDDGVKLMNMINPDKTEVRANKKEDFKYLEQYYPIDMESFNSNKEN 597
   |||||
Db 559 DEGTYSQVFSISGDDGVKLMNMINPDKTEVRANKKEDFKYLEQYYPIDMESFNSNKEN 618
   |||||
QY 598 VGDEKEIDPFKAPDIDKELKXEDIIIVPAGSTWSPRIDLLKPDVSAPEGKNIKSTLTNVIN 657
   |||||
Db 619 VGDEKEIDPFKAPDIDKELKXEDIIIVPAGSTWSPRIDLLKPDVSAPEGKNIKSTLTNVIN 678
   |||||
QY 658 GKSTYGY 664
   |||||
Db 679 GKSTYGY 685

```

RESULT 4  
AAM61246  
ID AAM61246 standard; Protein; 1007 AA.

AC AAM61246;  
DT 02-OCT-1998 (first entry)

DE Streptococcus pneumoniae SPI22 protein.  
KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
KW detection; pneumonia; otitis media; meningitis.

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX MO9818930-A2.

XX 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19422.

XX 31-OCT-1996; 96US-0029960.

PA (HDM-A-) HDMAN GENOME SCT INC.

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

DR MPI; 1998-272224/24.

XX N-PSDB; AAV27431.

PT Nucleic acid encoding antigenic peptide(s) from Streptococcus

XX pneumoniae - or their epitope-containing fragments, useful in

PT protective or therapeutic vaccines, and for diagnosis

XX Claim 11; Page 92-93; 118pp; English.

CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.

XX Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 19; Length 1007;  
Best Local Similarity 100.0%; Pred. No. 7e-72;  
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 38 ETSODPFEKKTAVIKEKEVVSKNPVIDNNTSNEAKIKEENSKSGDYTDSFVNNTEN 97
   |||||
Db 1 ETSODPFEKKTAVIKEKEVVSKNPVIDNNTSNEAKIKEENSKSGDYTDSFVNNTEN 60
   |||||
QY 98 PKKEDKVVYIAEFKDESEGEKAIKELISLAKTKVLTYYDRIFNGSAIETTPDMLDKIKOI 157
   |||||
Db 61 PKKEDKVVYIAEFKDESEGEKAIKELISLAKTKVLTYYDRIFNGSAIETTPDMLDKIKOI 120
   |||||
QY 158 EGISSEVERAKVQPMNNHARKEIGVEAIDYLSKINAPFGKNPDGAGMVISNIDGTGDR 217
   |||||
Db 121 EGISSEVERAKVQPMNNHARKEIGVEAIDYLSKINAPFGKNPDGAGMVISNIDGTGDR 180
   |||||
QY 218 HKAMRIDDDAKASMRPKEDLKGTDKNYMLSDKIPHAFFVYNGKITVEKYDDGRDYDP 277
   |||||
Db 181 HKAMRIDDDAKASMRPKEDLKGTDKNYMLSDKIPHAFFVYNGKITVEKYDDGRDYDP 240
   |||||
QY 278 HGMHLAGIILAGNDTEODIYNFNGIDGIAFPAQAIFSYKMTSDAGSGFAGDETMHAIEDSI 337
   |||||
Db 241 HGMHLAGIILAGNDTEODIYNFNGIDGIAFPAQAIFSYKMTSDAGSGFAGDETMHAIEDSI 300
   |||||
QY 338 KKNVDVSVSSGFTGTGLVGEKYMCAIRALRKAGIPMVVATGNYATSASSSSMDLVANNH 397
   |||||
Db 301 KKNVDVSVSSGFTGTGLVGEKYMCAIRALRKAGIPMVVATGNYATSASSSSMDLVANNH 360
   |||||
QY 398 LKMTDTGNVTRTAHEDAIIVASAKNQTVEFDKVNIGSEFKYRNIGAFPEKSKITTNED 457
   |||||
Db 361 LKMTDTGNVTRTAHEDAIIVASAKNQTVEFDKVNIGSEFKYRNIGAFPEKSKITTNED 420
   |||||
QY 458 GTKAPSKLKFVYIGKQDODLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIVNVN 517
   |||||
Db 421 GTKAPSKLKFVYIGKQDODLIGDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIVNVN 480
   |||||
QY 518 TNNVYNRDNTLPAWGADSTGKSOVFSISGDDGVKLMNMINPDKTEVRANKKEDFK 577
   |||||
Db 481 TNNVYNRDNTLPAWGADSTGKSOVFSISGDDGVKLMNMINPDKTEVRANKKEDFK 540
   |||||
QY 578 DKLEQYYPIDMESFNSKNENVDGEKEIDPFKAPDIDKELKXEDIIIVPAGSTWSPRIDLL 637
   |||||
Db 541 DKLEQYYPIDMESFNSKNENVDGEKEIDPFKAPDIDKELKXEDIIIVPAGSTWSPRIDLL 600
   |||||
QY 638 LKPDVSAPEGKNIKSTLTNVINGKSTYGY 664
   |||||
Db 601 LKPDVSAPEGKNIKSTLTNVINGKSTYGY 627

```

RESULT 5  
ABP54664  
ID ABP54664 standard; Protein; 1007 AA.

XX ABP54664;

DT 04-SEP-2002 (first entry)

XX S. pneumoniae SPI22 protein sequence SEQ ID NO:216.

XX Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

XX antibacterial; Streptococcal infection; detection.

OS Streptococcus pneumoniae.

XX US2002061545-A1.

XX 23-MAY-2002.

PF 22-JAN-2001; 2001US-0765272.

XX 30-OCT-1997; 97US-0961083.

PA (CHOI/) CHOI G H.  
PA (KUNS/) KUNSCH C A.  
PA (BARA/) BARASH S C.  
PA (DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.  
 PA (FANN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 XX  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,  
 PI Rosen CA;  
 DR WPI; 2002-479261/51.  
 DR N-PSDB; ABQ84899.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 50; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 XX  
 SQ Sequence 1007 AA;  
 Query Match 94.8%; Score 4171; DB 23; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 7e-72;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 ETSODPFEKKTAVKEKEVSKNPVINDNTSNEAKIKEENSKNSQGGYDTSFVNKNTEN 97  
 DB 1 ETSODPFEKKTAVKEKEVSKNPVINDNTSNEAKIKEENSKNSQGGYDTSFVNKNTEN 60  
 QY 98 PKKEDKVVYIAEFKXSEGEKAIKELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOI 157  
 DB 61 PKKEDKVVYIAEFKXSEGEKAIKELSLKNTKVLTYTDRIFNGSAIETTPDNLDKIKOI 120  
 QY 158 EGISSVRAQKVQPMNHARKEIVEEALDYKINAPFGKNPFGKGVTSNIDTGTDR 217  
 DB 121 EGISSVRAQKVQPMNHARKEIVEEALDYKINAPFGKNPFGKGVTSNIDTGTDR 180  
 QY 218 HKAMRIDDDAKASRFKKEDLKTGTDKNYMLSDKI.PHAFNYNGGKITVEKTDGGRDYFDP 277  
 DB 181 HKAMRIDDDAKASRFKKEDLKTGTDKNYMLSDKI.PHAFNYNGGKITVEKTDGGRDYFDP 240  
 QY 278 HGMHIAGILAGNDTEODIKNFNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSI 337  
 DB 241 HGMHIAGILAGNDTEODIKNFNGIDGIAIPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSI 300  
 QY 338 KANVDVSVSSGFGTGLVEKTYQAIRALKAGIPMVVATGNTATSASSSSMDLVANNH 397  
 DB 301 KANVDVSVSSGFGTGLVEKTYQAIRALKAGIPMVVATGNTATSASSSSMDLVANNH 360  
 QY 398 LKMTDGNMVRTAHEDAIAVASAKQTVBERDKYNIGSESPKYNTIAFPDKSKITTNEED 457  
 DB 361 LKMTDGNMVRTAHEDAIAVASAKQTVBERDKYNIGSESPKYNTIAFPDKSKITTNEED 420  
 QY 458 GTKAPSLKLFVYIGKGDDDLIGDLRGKIAVMMDRIYTKDLKNAFKKAMDGARAIVVN 517  
 DB 421 GTKAPSLKLFVYIGKGDDDLIGDLRGKIAVMMDRIYTKDLKNAFKKAMDGARAIVVN 480  
 QY 518 TVNYNRDNTLPLAMGYEADGEGTSQVFSISGDDGYKLANMINPDKKTEYKRNKEDFK 577  
 DB 481 TVNYNRDNTLPLAMGYEADGEGTSQVFSISGDDGYKLANMINPDKKTEYKRNKEDFK 540  
 QY 578 DKLEQYYPIMESFNSKPNVGEDEKIDPKFAPPTDDELVEKEDIIVAGSTWGPRIDLL 637  
 DB 541 DKLEQYYPIMESFNSKPNVGEDEKIDPKFAPPTDDELVEKEDIIVAGSTWGPRIDLL 600  
 QY 638 LKPDVSAFGKNIKSTLNVINGKSTYGY 664

DB 601 LKPDVSAFGKNIKSTLNVINGKSTYGY 627  
 RESULT 6  
 ID ABB98574 standard; protein; 2835 AA.  
 XX ABB98574;  
 AC  
 XX 14-JAN-2003 (first entry)  
 DT  
 XX Dextran saccharase, DSRE.  
 XX Dextran saccharase; enzyme; cytosolic; dermatological; antiseborrheic;  
 KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;  
 KW microflora regulation; intestinal transit; mineral assimilation;  
 KW colon cancer; acne; dandruff; body odour.  
 XX Leuconostoc mesenteroides NRRL B-1299.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..40  
 FT /label= Signal\_peptide  
 FT 41..2835  
 FT /label= Mature\_protein  
 FT 423..439  
 FT /label= SEQ\_ID\_6  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT /label= SEQ\_ID\_7  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 519..539  
 FT /label= SEQ\_ID\_8  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 560..571  
 FT /label= SEQ\_ID\_9  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 631..645  
 FT /label= SEQ\_ID\_10  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 1014..1021  
 FT /label= SEQ\_ID\_11  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2120..2138  
 FT /label= SEQ\_ID\_12  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2161..2184  
 FT /label= SEQ\_ID\_13  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2202..2214  
 FT /label= SEQ\_ID\_14  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2243..2250  
 FT /label= SEQ\_ID\_15  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2315..2332  
 FT /label= SEQ\_ID\_16  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2689..2696  
 FT /label= SEQ\_ID\_17  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 1981..1112  
 FT /note= "Catalytic domain, SEQ ID 1. This sequence is specifically claimed in Claim 4"  
 PR 16-MAR-2001; 2001FR-0003631.  
 PR

XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Beccomet SAM, Renaud SMC, Willemot RML, Monsan PEF;  
 XX WPI, 2002-715213/78.  
 XX N-PSDB; AB080961, AB080962.  
 DR New glycosyl transferase enzymes, containing glucan bonding and  
 XX catalytic domains and producing alpha-(1-2) branched dextrans, useful  
 PT in probiotic, pharmaceutical or cosmetic compositions -  
 XX Claim 6, Page 65-74, 82pp; French.  
 XX The present sequence is a novel dextran saccharase, DSR-E, from  
 CC Leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl  
 CC transferase activity suitable for producing dextrans having alpha(1-2)  
 CC branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,  
 CC alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbiptanoside or  
 CC alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in  
 CC prebiotic, pharmaceutical or cosmetic compositions. The dextrans and  
 CC related compounds having alpha(1-2) bonds, produced using DSR-E, may be  
 CC involved in signalling/cellular recognition processes in vivo  
 CC (specifically in regulation of microflora in the intestines or on the  
 CC skin); and are potentially useful for improving intestinal transit,  
 CC increasing assimilation of minerals (e.g. calcium and/or magnesium),  
 CC preventing cancer of the colon and combating skin problems such as acne,  
 CC dandruff and body odour.  
 CC  
 XX Sequence 2835 AA;  
 SQ  
 Query Match 35.2%; Score 1549; DB 23; Length 2835;  
 Best Local Similarity 24.7%; Pred. No. 5.2e-16;  
 Matches 289; Conservative 150; Mismatches 149; Indels 581; Gaps 219;

DB 1036 GSGKDYDSVINNTLY-DSRTVGGGEYQEXFGGLFLDQLKDYPSLETRQI-STNOPMN 1093  
 QY 341 VDVVSV---SSG-FTGT---GLVGEKY---W---Q-AIRALRKAG-IP-MV---VAT 378  
 DB 1094 PD-VKIKEMSAKYPNSNIQG-RGANYVLKDMATNQYFNVS--DNGFLKQLLGKSTST 1149  
 QY 379 G---N-----YATS---A-S-----SSWDLVANNHLKXTDTG--NVTRTAAHED--- 414  
 DB 1150 GFITTEKGTSPYSTSGVQAKMDTFIDQGTWYVFDNAGY-ML-TGKQNI-----H-DKQVY 1201  
 QY 415 ---ALAVASA---KQVVEPKVNIIGESFK---Y---RNGA-PEDKS-K-I--T-- 453  
 DB 1202 FLPNGLVDDAYLFDNQ--EF-YNNKAGEQVNNQYQDSQONQWYFFENGRAALGLTEV 1258  
 QY 454 TNEDEGTAKPSKLF---VYI-G---KGQ-DQ---D-----LI---G-L-D-----L- 483  
 DB 1259 PNAADGTHV-TQY-FDANGQIKGTALKDQNNQRLRYDEATGNVNVSWGLADKSWLYLN 1316  
 QY 484 -RCKIAVM-DR-LYTKDLNA--FKKAMD---KGARAIMVNTVN---YNNRD----- 525  
 DB 1317 AQG-VAVTGNOXI---DGE-EYVF-NA-DQKQYKGN-AI-IDN--NGDQRYDGDGKGMV 1365  
 QY 526 -N-WTELLPAMG---YEADG---T-----KSQV-FSISGDDG--VK-----L---W--- 557  
 DB 1366 VNSWGLPD-GSWLYLNDKGIAYTGROVINNVNF--FGNDKQIKDAFKLLDGSWVYL 1422  
 QY 558 -NM-----INP-----DK-----K-----TEV--KR-----N-----NK--E--D- 575  
 DB 1423 DDKGLITTGAKVINGAMFFDQDGHQIKGDASTDANGKHYYDKNDCHLYTNSWGLPDG 1482  
 QY 576 ---F-K-DKL-EQYYPIDMS-F-----NSNKP--NVGDEKEI---DFKEAPD 611  
 DB 1483 SWLYTEEGQDAVYQGRV-IDGKTRVFEDEGKQIKNSLKTLAN-GD-K-IYLDGD-GVA-A 1536  
 QY 612 T-----DKELY-KED---I---VPA--GSTW-----GP-RID---LLKLPD 641  
 DB 1537 TGLQHVGDKIMYFDEDEGQVNGKFSVAKDG--SWYLLNDGVAAGPSSYINGOSLYPDQD 1594  
 QY 642 VSAPGNIKSTLN-VIN--GKST--Y-GY 664  
 DB 1595 ---GKQVK--YNEVNSDQ--TTNYTTGL 1616

RESULT 7  
 AAR08221  
 ID AAR08221 standard; protein; 1475 AA.  
 XX  
 AC AAR08221;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 06-MAR-1991 (first entry)  
 XX  
 DE Recombinant alpha amylase pullulanase enzyme.  
 XX  
 KW Starch; pullulan; saccharification; debranching; liquefaction;  
 KW hydrolysis.  
 OS Clostridium thermohydrosulphuricum DSM 3783.  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..31  
 FT /label= signal peptide  
 XX  
 PN EP402092-A.  
 PD 12-DEC-1990.  
 XX  
 PD 05-JUN-1990; 90BP-0306088.  
 XX  
 PR 05-JUN-1989; 89US-0361368.  
 XX  
 PA (ALKO-) ALKO LTD.  
 PA (ARCO-) ARCO OY AB.

```

XX Melanem H, Palohelmo M;
XX WPI: 1990-370103/50.
DR P-PSDB: AAR08221.
XX
PT Heat stable enzyme with both alpha-amylase and pullulanase
PT activities - prep'd. by expressing Clostridium thermohydrosulphuricum
PT DNA in host cells, useful in hydrolysis of starch, etc.
XX
PS Claim 6; Fig 5; 127bp; English.
XX
CC The recombinant enzyme was produced by cloning the gene encoding it
CC from a C. thermohydrosulphuricum genomic library and using it to
CC express the enzyme in a host cell e.g. E.coli which is easier to
CC cultivate and has less complex nutritional requirements than C.
CC thermohydrosulphuricum ( which is an obligate anaerobic thermophilic
CC organism). The enzyme is used to hydrolyse starch, amylose or
CC pullulan, opt. in conjunction with a glucogenic or maltogenic enzyme.
CC It has an optimum temp. range of 80-85 deg.C, i.e. 5 deg. lower than
CC that of the native enzyme, but the heat stability is the same. At
CC least 10 differently sized polypeptides having the same enzymic
CC activities were produced, of Mr 100 000 - 165 000.
CC (updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 1475 AA;
Query Match 35.0%; Score 1541; DB 11; Length 1475;
Best Local Similarity 23.4%; Pred. No. 3.9e-17;
Matches 299; Conservative 169; Mismatches 126; Indels 656; Gaps 242;
QY 2 P-----VV-----LADTS-----SSED-----AL--N-----IS-----DKE-KVAENK 29
DB 37 PAIANYGNFQSKGLGSDMNINS-DKTIIMYKNGFYEFTTPVALPAGDYEYKALN----- 92
QY 30 H--E-----NI--HSAETSOD-----F-KEKTAIVKEKEVSK-NPVI--DN----- 65
DB 93 HSWGGGVPSSGNSLFSH-LD-S-DSVVTFFYNNYNTSITDS--TKYTP-IPEDKLPRL 144
QY 66 -NTSNEAKI-----KEENSNGKS-OGDYTDSFVNKOT-E--N-PKKEDEVVYIAEPK- 111
DB 145 VGT-IGPA-IGAGDMKPESTAIIR-DY--KF-N-NVEYETIANVPG-N--Y-EFKV 191
QY 112 -----D-----KESGE-----KAIKELS-SLKNTKV-LYTYDRIF-NGSAIET-----T-PD 149
DB 192 TLGSPWINTNGLENGGNPI-PLNVAY-DTKIRFY-YDSVSH-----IMDYNEPLTGPD 245
QY 150 N---LDKIK-----QIE-G-I-S--SV-ERAKVQPMNTHARKEIGVEEA-IDY--- 188
DB 246 NNIYYDLRHDTHDPFRSPFGAIKTGDVTLR---IQ-AKNH---DI--ESAKISYWD 296
QY 189 L-KS-IAP-F--GKNFGR---GMV--ISNIDTCTDRHKAMR-----D----- 224
DB 297 IKTIRIEVPMRIQSPDGKYEWEVLS-FD-----HPT-RIMWYFILDKGTAYYG 348
QY 225 D-D-----AKASMRFKEDLKTG--DKNY-----WL--SDK-I-PAFVNYNGSKT-- 264
DB 349 DNBEQLGCVGKAT-DTENKQFELTVYDKNLDTPMGMGSVVYQFPBRF--FNDSNDH 405
QY 265 VEKYDGRDYFDP-HGMHIAIGI-L-AGND-TE--Q-D-I-KN-F-----NGID----- 302
DB 406 LKKYSRG---FDPVE-YHSNMYELPDNPNDKNKLGYGDGWSNDPFGDLKIDDKLDY 461
QY 303 -----GIA-----PAAQIF-S-YKM-Y--SDAGSG---FAGDETF-----H-AIE- 334
DB 462 LKSLGISIVIYNP---IFQSPSNHRYDTTD-NTYIDELGLSTFKKLMEDAHAKAKVI 517
QY 335 -DSI-KANVD--VVSVES-G-F--TGT-G-L-V--G-E-K-Y--WOAIR-----ALRKA- 370
DB 518 LDGVFNHTSDSIY-FDRYGYKLTGVLGAQAKQDQSSPYGDMYEIKPDQTY-EGW 575
QY 371 -G---IMVV--ATGN-YATSSASSSW-DLVANNH-L-K-K-MT-DTG--NV-T--R-- 408

```

```

DB 576 WGFDSLIP-VIRQINGSEYV-----KSMADFTINNPNPAISKYTLNPD-GDRNAGADGKRLD 629
QY 409 TA---AHEDA-I-----AVASAK-NO-TV-E-F-D-KVNIAGESP---K-Y--NN-IGAF- 446
DB 630 VANEVAH-DEWVHFRGALINTVKNAPMVAENMNDASLDLGDSPNSVMNVLFFNNVADF1 688
QY 447 FDYS-----K-----I-----TT-N---ED-----G-TKA 461
DB 689 FDKSPDDGNVHNPNINAAKLIDRLMSIYERYPLVPEYSTWNLGSHDTWRILTVFGYNSA 748
QY 462 -P---S-----K-LKF--V-----Y-----I--G-K-GQ-----QOD--LI--G--L 481
DB 749 DENQNSQAKDLAVKULKALAILQMGYPCGMSIYVGDAGSGGKDPDNRRTTPWGEDT 808
QY 482 DLR---GKIA-VN--DRIV-TKDLK---N---AF-KK---AMD-KG-----ARAIMV 516
DB 809 DLDTFFKRVNINENOVLTGDLLETLVANGDVYAGRRIRINKDTPGSKPPPSVALVI 868
QY 517 N-----TV--NYNRD--NMTLPAM-G-YE-AD-----E-GTK-SQVFSIS--G-- 550
DB 869 NKGDQKQVSIDTTFKFRDGVAFPD--ALSGKTYVDGKIVGVSGMDGAIL-ISPQGN 925
QY 551 -----D-----DG-VKL-W-----N-----MI--N-----PDKK 565
DB 926 LTPAQPTDLKAVSGNGKDLSSVVDKAVSYNIYRSTVKGLEYEKIASNVTOITYTD-- 983
QY 566 TEV-K--R-----N-NKEDFKDLKQY--YPI-----DKMESFNS-----NKP----- 596
DB 984 TBTTLNGLRYVAVTAANDNGBESALSNEVAYAPFPGAGMNNQVTHYIGNNPVEVY 1043
QY 597 -NV--G--DEK--E--ID--FKPAPDT-D--KELY-K--E--D--I-----I 622
DB 1044 AEVMAQGLTDKPGQGENMIQLGYRYIGDTGVDAVYNAVYKVEGVEISKDMTMDAQVY 1103
QY 623 -----VP--AGS-----TS-----W-----GP--RID---LLKP--DV- 642
DB 1104 GDSGNNDKYMAKFVPMVGVYIMRFSNGQHDWYTYKGPDKGTBAKQFTVPSNDVE 1163
QY 643 -SA-----PGKNIKS---TLN-----V-INGKSTY--GY 664
DB 1164 TPTAPVLQDPG--IESRVTLMNSPSADVAIRG---YEYI 1199

```

RESULT 8  
AAR95016  
ID AAR95016 standard; Protein; 2710 AA.

XX AAR95016;  
XX  
XX  
DT 08-JUL-1996 (first entry)  
XX  
DE C. difficile toxin A.  
XX  
KM Toxin A; cytotoxin; enterotoxin; fusion protein; antitoxin;  
XX vaccine; diarrhoea; diagnosis; therapy.  
OS Clostridium difficile VPI strain 10463 (ATCC 10463).  
XX  
PN W09612802-A1.  
PD  
PD 02-MAY-1996.  
XX  
PF 23-OCT-1995; 95MO-US13737.  
XX  
XX 07-JUN-1995; 95US-0480604.  
PR 24-OCT-1994; 94US-0329154.  
PR 16-MAR-1995; 95US-0405496.  
PR 14-APR-1995; 95US-0422711.  
PA (OPHT-) OPHIDIAN PHARM INC.  
XX  
XX Flicra JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;  
PI Williams JA;



XX MP1: 1996-230603/23.  
DR N-ESDB: AAT29248.  
XX  
PT Fusion proteins comprising non-toxin protein and part of toxin -  
PT useful to form anti-toxins against Clostridium botulinum type A, and  
PT C. difficile type toxins, and to treat C. difficile intoxication,  
PT partic. diarrhoea  
XX  
PS Claim 63; Page 290-302; 434pp; English.  
XX  
CC Clostridium difficile VPI strain 10463 toxin A (AA95016), the  
CC product of the toxin A gene (AAT29248), is a potent cytotoxin that  
CC plays a direct role in damaging gastrointestinal tissues and is  
CC associated with diarrhoeic disease. It can be obtd. by expression in  
CC transformed E. coli hosts of portions of DNA that together cover the  
CC entire toxin A gene. Toxin A, and portions of it (see also  
CC AA95014-15 and AA95017), pref. expressed as fusions to polynucleotide  
CC affinity tags or maltose binding protein, are used to raise avian  
CC antibodies useful as antitoxins or diagnostics, and in vaccine prodn.  
XX  
SQ Sequence 2710 AA;

Query Match	34.7%;	Score 1525;	DB 17;	Length 2710;
Best Local Similarity	23.8%;	Pred. No. 1.3e-15;		
Matches 300;	Conservative 165;	Mismatches 137;	Indels 659;	Gaps 258

```

09      1  YPVVLA-D-----TS--S-S-E-----D-----ALNISDDE---K-----VAENKEK---H 30
      2  :::::
1426  12  YSLSLGDKNYLISLNSNTTEKINTLGLDSKNIAAYTD--ESNNKYFGAALSKTSQCSIH 148
      13  :::::

```

```
09 -----H--SAME-TSQDF-KEKTRAIVIKEKV--VS-KMPVIDNNT-----67
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1485 YKKDSKNLIEFYNDSTLEEFNSKDFAE-DINVFPMKDINTITGK-YVDNDNTDKSIDFSI 1544
```

```

QY      68  ---S-NEEAKIKEENS---NKS-QGDYDTDSFV-NK---NTENPKK-EDKV-VY-I-A- 108
      1543 SLVSKN-QVKV---NGLYALNEVSYSYLD-FVKSGDHNNHTSNFNNLLFDLNIISFKLPGF 159
Db

```

QY 109 E--F-KDKE-S--GEKA----IKELSLAKTK-V-LY-----T-YDR--IF--NG--SA 143  
Db 1598 ENINFEVIDIKKFTLLVG-KTNGYV-BF-ICNNKNKIDIFYGWEKWTSSSKSTIFSGGRNV 165

```

Qy      144 IET--TPDN-LDKIKO-I----E-----GISSVER-AOKV--QP-----MNNHARKEIGVE- 183
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      1655 VEPYNPDPGDED-ISTSLDSEYEPLYGI---DRYINKVLIAAPDLTSLIN-----INTNY 170
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Qy	184	EAIDYLKST---	NAPFGKN	FDGRGV	ISNID	---	T-GTD	----	Y-----	216
Db	1706	YSNEYEPEITLVN-P---	HNF	HKVNI	NLDSS	FEYKSTGSD	FLIVRYLEESNK	175		

Oy	217	----	RHK-----	AMRID-DDAKA-S----	MR-FK-----	K-E----	D-LKG----	TD	242
Db	1759	110KRIKIGILSTNOSFNKNSIDFQDKKISLGYATMSNFSEFSENSEIILDBDHU--	EPKIID	181					

Qy	243	-KNYWL-SD---	K----	I-----	PHAFNY---	Y--	NGGX----	I-T-VE--	KYD-	269
		: : : : : :	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	: : : : :	
Db	1818	NKTYTYVDEDEKRI	NYKGI	ININNSI	FYEDPTREIN	YVNGCMOT	NGKRYVYENI	NTGAAT	TSYKI	1877

QY    270 -DGD-YF--DPHG-MHAGILAGND-----T-----E-QDI---KNF--NG--   300  
              |      |      |      |      |      |      |      |      |      |  
DB    1878 TNGCFEYFNND--GVNCL-VEUEKCDCEFEYADAMTONDDREGLITVNSC-FLETKGKY 1923

[illegible]

```

0y      322  -G-----FAGD-ET--MF--H-AIEDSIKH---NVD--WSVSSG-F-TGTG--LVGEK--Y 360
          | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0s      1000  MCGGVVF--LMPVPLVPGVYVVI--DC--KAPAPGQVPLVZ--:--CUTGQVQV--GPPNPPV--PAMVY 304

```

QY            361 -----MOATPALR-K-----AGIPMTV-----A-TG-----N-----Y-AT-SA-SSSSW--         390  
             :|||         |         :|||         :         |         :|||

Db	2046	NNNIEGAI-VYOGSKFLTLNG-KKYYFDNNNSKAVTGQITIDSKKYYFNTNTAAEAATGMQT	2103
Qy	391	-D---LVAN-NHLKM-----T-DTG-----NVTRTA-A-----H-----ED---AI	416
Db	2104	IDGKKYFNTNTAAEAATGMQITID-GKKYYFN-TNTAASGTYYTINGKHHYFNTDGMQI	2161
Qy	417	AVASAKQOTVE-F-----DKVNIIGSESPKRYN-----I-GA--FF--DKSK-IT-----TN	455
Db	2162	GVFPGRN-GREYFAPANTDANNIEGAI LYONBELTINGKKYYGSD-SKAVTQMRILNN	2219
Qy	456	EDGTR--AP-SKLKV-Y--IGKGQODDILGLD-LR-GKIAVMDR-IYYTD-----	497
Db	2220	K---KYYFNPNNAIAIHLCTINN--DKYYFSYDGILONGYITI-ERNFNY-FDANNESK	2272
Qy	498	-----LK--YAFK---KA--MD---KGARAIMV-N---TVN---RY--NRD-----NW-	527
Db	2273	MVTGVFGPGRGFEYFAPANTNNNIEG-QAI-VYONFELTINGKKYYFDN-DSKAVTGWQ	2329
Qy	528	T-----E-----L-----PAM-GYEA-DEG-----T-K--S--Q-V-----F---S-I-S	549
Db	2330	TIDGKKYYFNLTAAEAATGMQITID-GKKYYFNLTAAEAATGMQITIDGKKYYFNTNTITIAS	2388
Qy	550	-G--D-DGVK-LW-N-----M-I-----N-----P--DKKT-E--V-KRN-----N-	572
Db	2389	TGTYTSING-KHFYFNTDGMQIGVFKPGRGHEYFAPANTDANNIEGAI LYONKFLTLNG	2447
Qy	573	KE-DP-KD-K---L-----EQQY-----PID-ME-SFSPN-----KP--NV-598	
Db	2448	KKYYFGSDSKAVTGRTITIDGKKYYFNTNTAAVATGMQITINGKKYYFNTNTNISIASTGYITI	2507
Qy	599	-G-----DE-KEI-----D-FK-PAP-DTDK---E-----LYKEDIIVPAG-	626
Db	2508	SGKHFYFNTDGMQIGVFKPGRGDEYFAPANTDANNIEGAI RYONRFLVLIHDIYYFGN	2567
Qy	627	-S---TSMGPRID---LLKPDVASPKN-IKSTL---N-VI-NG-----K-STYG--	663
Db	2568	NSKATGMVT-IDGNRYFYFEBNTAM--GANGYK-TIDMKNFYFRGLPOLQIGVFKGSN-GFE	2623
Qy	664	Y 664	
Db	2624	Y 2624	

RESULT 9	
AAW68387	
ID	AAW68387 standard; Protein; 2710 AA

AC	AAW68387;
XX	
DT	07-DEC-1998 (first entry)

DE	Clostridium difficile toxin A.
XX	
KW	Antitoxin; vaccine; cytotoxin; toxin A; intoxication; immunogen.

XX Clostridium difficile.  
OS  
XX

05-MAR-1998.

PR	28-AUG-1996;	96US-0704159.
XX		
XX		

XX  
PI  
XX  
Thalley BS, Williams JA,  
XX

DR WPI; 1998-230234/20  
DR N-PSDB; AAV30560.  
XX



Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin

Example 15; Page 220-230; 428pp; English.

This is the amino acid sequence of Clostridium difficile toxin A, deduced from the coding region (see AA050560) of the toxin A gene. Toxin A is a potent cytotoxin that plays a direct role in damaging gastrointestinal tissues. Severe cases of C. difficile intoxication result in pseudomembranous colitis. This would be prevented by neutralising the effects of toxin A in the gastrointestinal tract. Examples are provided of the production of recombinant C. difficile toxin A in host cells and of the in vivo neutralisation of toxin A by antibodies against recombinant toxin A polypeptides. The invention specifically relates to recombinant proteins derived from Clostridium botulinum toxins (see AA068385-400) and their use as immunogens for the production of vaccines and antitoxins.

Sequence 2710 AA;

Query Match 34.7%; Score 1525; DB 19; Length 2710;  
Best Local Similarity 23.8%; Pred. No. 1,3e-15;  
Matches 300; Conservative 165; Mismatches 137; Indels 659; Gaps 258;

1 YPVVLA-D---TS--S-S-E-----D---ALNISDKE---K---VAENKEK---H 30  
1426 YSLLSDSKVYLNLSNTEIKNTLGLDSKNIAVNTD-ESNNKYGAISKTSQKSIH 1484  
31 ----ENI---H---SAME-TSQDF-KEKTAIVIKEV--VS-KNPVIDNNT----- 67  
1485 YKSDSKNILEFYNSTLEFNSKDFIAE-DIVFMKDDINTITGK-YVVDNNTKSIDPSI 1542  
68 ----S-S-NEAKIKEKNS---NKS-OGDYTDSFY-NK---NTENKPK-EDKV-VY-I-A- 108  
1543 SLVSKN-QVKV---NGLYNESVYSYLD-FVKNSDGHNTSMNLFNDNISFWKLGFG 1597  
109 E---F-KDKE-S---GEKA---IKELSLKNTK-V-LY-----T-YDR--IF--NG--SA 143  
1598 ENINFDVDKFTPLVG-KTNLGYV-EF-ICDNNKIDIDYFGWKTSKSTIFSGNGRNV 1654  
144 IET--TPDN-LDKIKO-I---E---GISSEYER-AQRY-OP-----MMNRARKEIGVE- 183  
1655 VERLYNDTGED-ISTSLDSFYELGYI---DRYINVLAPLDYLTSLN-----INTNY 1705  
184 EALDYLSI---NAPFGKN-FDGRGV-ISNID-----T-GTD-----Y----- 216  
1706 YSNEYEPIIVLN-P---NTF--HKYVNI-NLDSSEFEYKWKSTEGSDFILVRYLESNKK 1758  
217 ----RHK-----AMRID-DPAKA-S---MR-PK---K-E---D-LKG---TD 242  
1759 ILQKIRIKGLISNQSFRKSIDPKDIKLSLGYTMSNFSNSENELDRDHL-GFKIID 1817  
243 -KNYWL-SD---K---I-----PHAFNY--Y---NGKR---I-T-VS--KYD- 269  
1818 NKTYVDEDSGLVGLININSLFYFPIEBNLVTGMQTINGKXYFDMTGALTSYKI 1877  
270 -DGRD-YF--DPHG-MHAGIAGND-----T-----E-ODI---KNF---NG-- 300  
1878 INGHGFYFND--GVML-GVFKGPDGEFAPANTNNNIEGQAIYQSK-FLLNGKK 1933  
301 --ID-----G---I-----APN---A---Q-I--PSYKNV--SD-A---G--S- 321  
1934 YFPDNNKAVTGMRIINNEKYPNPNAIAVGLQVINDNKT--YFNPDRILISKGMQTV 1991  
322 -G---FAGD-ET--MF--H-AIEDSIRK--NVD-VVSVSSG-F-TGTG--LVGER--Y 360  
1992 NGSRYF--DTDLIAFNGYKTI-DG-KHFYFSDCVYKI--GVFSINSNGEYFAPANTY 2045  
361 ----WQAIRLR-K---AGIPNV-----A-TG---N---Y-AT-SA-SSSGW-- 390  
2046 NNNIEGQAI-VYQSKFPLTNG-KKYYPDNNKAVTGLQITDSKKYFNTNTAEAAVMQOT 2103

QY 391 -D---LVAN-NHUKM-----T-DTG-----NTRTA-A-----H-----ED---AI 416  
DB 2104 IDGKKYFNTNTAEATGMQITD-GKKYFN-INTNASTGYTILINKHFFNTDGMQI 2161  
QY 417 AVASAKQVTE-F-----DKVNIGSESFKYRN-----I-GA--FF--DKSK-IT-----TN 455  
DB 2162 GVFKGPN-GFEYFAPANTDANNIEGQAILYQNEFLTLNGKKYFGSD-SKAVTGMRIINN 2219  
QY 456 EDGTK--AP-SLKPV-Y---IKGQDQDQIGD--LR-GKIAMD--IYTD----- 497  
DB 2220 K---KYFBNNAIAAHLCCTINN--DKYFSYDGLQNGYITI-ERNFY-PDANNESK 2272  
QY 498 ----LK--NAFK---KA--MD---KGARIMV-N---TVN---YY--NRD-----NW- 527  
DB 2273 MVTGVFGRPGFEPAPANTHNNNIGS-QAI-VYQKFLTLNKKYFIDN-DSKAVTGMQ 2329  
QY 528 T---E---L---PAM-GYEA-DEG-----T-K-S-Q-V-----F--S-I-S 549  
DB 2330 TIDGKKYFNLNTAEATGMQITD-GKKYFNLNTAEATGMQITD-GKKYFNTNFIAS 2388  
QY 550 -G--D-DGVK-LM-N---M-I-----N-----P---DKT-E--V-KRN-----N- 572  
DB 2389 TGYTSING-KHFYFNTDGMQIGVFKGPNGEYFAPANTDANNIEGQAILYQNKFTLNG 2447  
QY 573 KE-DF-KD-K---L-----EQQY-----PID-ME-SFNSN---KP--NV- 598  
DB 2448 KXYFSGDSKAVVGLRTIDGKKYFNTNTAVATGMQTINGKXYFNTNISTASTYTI 2507  
QY 599 -G-----DE-KEI-----D-PK-PAP-DTDK--E-----LYKEDIYVAG- 626  
DB 2508 SGKHFEYFNTDGMQIGVFKGPDGEFEPAPANTDANNIEGQAIRYQNFYLYLHDIYFEGN 2567  
QY 627 -S---TSWGRID---LLKPDVSAPGN-IKSTL--N-VI-NG-----K-STG-- 663  
DB 2568 NSKAATGWT-IDGNRYFEFPTAM-GANGYK-TIDNKYFPRNGLPQIGVFGSN-GFE 2623  
QY 664 Y 664  
DB 2624 Y 2624

RESULT 10  
ABB47329  
ID ABB47329 standard; Protein; 1530 AA.  
XX ABB47329;  
AC  
XX  
DT 05-FEB-2002 (first entry)  
XX  
DE Listeria monocytogenes protein #33.  
XX  
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease.  
OS Listeria monocytogenes.  
XX  
PN W0200177335-A2.  
XX  
PD 18-OCT-2001.  
XX  
PF 11-APR-2001; 2001MO-FR01118.  
XX  
PR 11-APR-2000; 2000FR-0004629.  
XX  
PA (INSP ) INST PASTEUR.  
PI Buchrieser C, Frangeul L, Couve E, Ruanjok C, Pahi H, Dehoux P;  
PI Duesenget O, Chetoui F, Nedjari H, Glaser P, Kunet F, Cossart P;  
PI Daniels U, Goebel W, Krefe U, Kuhn M, Ng B, Vazquez-Boland JA;  
PI Dominguez-Bernal G, Garrido-Garcia P, Pierrez-Martinez A, Amend A;  
PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L;  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

P1 Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;  
 P1 Rose M, Voss H;  
 XX WPI: 2002-010914/01.  
 XX  
 PT Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment  
 PT and prevention of *Listeria* and related bacterial infections, and  
 PT related polypeptides -  
 XX  
 PS Claim 6; SEQ ID No 34; 192pp; French.  
 XX  
 CC The present invention relates to the genome sequence of *Listeria*  
 CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of  
 CC it are useful for selecting probes and primers for detecting genes in *L.*  
 CC *monocytogenes* and related organisms, and for studying genetic  
 CC polymorphisms and other genomes. The present sequence is a protein  
 CC encoded by the genome sequence of the present invention. Proteins  
 CC expressed from the genome sequence are useful for raising specific  
 CC antibodies, identification of *L. monocytogenes* and related organisms, and  
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
 CC B12. The genome sequence and proteins encoded by it are also useful for  
 CC selecting compounds that regulate gene expression and cell replication  
 CC and modulate *L. monocytogenes*-related diseases. In addition, the genome  
 CC sequence and proteins encoded by it are useful in pharmaceutical and  
 CC vaccines compositions for the treatment or prevention of infections by *L.*  
 CC *monocytogenes* and related organisms.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIRO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1530 AA;  
 Query Match 34.4%; Score 1511; DB 23; Length 1530;  
 Best Local Similarity 24.6%; Pred. No. 1.9e-16;  
 Matches 302; Conservative 162; Mismatches 127; Indels 635; Gaps 248;  
 QY 1 YPV-----V-LA-DTSSS-E---DAL-N-ISDKE-KVAENKE--KHENHSAME- 38  
 DB 163 FPISENDEEPTIKIKPDTNAPFEKKGSTVQNAQIWEIKI--NQLATHENA-SV-ED 218  
 QY 39 ---TSQ---D-FKE--KKTAVIK-EKEVSK-NP---V-IDNNTENEAKIKE--EN 78  
 DB 219 FPGNQOKILFDSF-EVHEMNMHIDGTVEIGAKVPESEYVTKI--NNTSG-DTGFLPTREN 275  
 QY 79 SNKS--OGDY---TDSFVNK---NTE--N--PKKE--D-KVY-----YIAE-----EKD 112  
 DB 276 PIDSAVQTYTQTEITD-F-NKATFSNTATFNSDETEEVMSDDVATPRDPITQKPLFLF-D 332  
 QY 113 KESGEKAI--K-ELS---SLKNTKVLV-TY-D-----RIFNG-SAI-----E-T-T 147  
 DB 333 TTSN-K-ISMVKEVNGHYALDNS-VWVDYTIDEGSPGIGIRIVAGDGIPLLPTEYIVT 389  
 QY 148 -----PDNLD--KI--KQI--EG---ISSVERAKQVPMN--NHAKKEIG-- 181  
 DB 390 DMGSAGFRIEIPD--DGYTYITFETIVEGETGMISNT--ASIIISPNDNEA--E-GDY 442  
 QY 182 -VB-----EA-IDYLSI--NAPF---GK-----NFDGGMVTSN--I-DTGTD--Y- 216  
 DB 443 FIPDPSPEGLID--KLIEN--FNAKTGLTWEIITINKDG-G-TLHNPVITDEPFGGLI 496  
 QY 217 RHKA-MRI-DDD-AK-ASMRPKE--DKGTDKNTW-----LS-D-----K--I 251  
 DB 497 FHPATLKIMDSBOVLDSADY--EVVPLNG--DSS-WKSGFOINFRDITGQHIVITKYQV 552  
 QY 252 -P--HAF--N-YUNGKI--TVB-----KY-D--DGRD-Y---FD-PHGMIAG-- 284  
 DB 553 NPSTHSGDNEYNNATIKTDTAEASDSDTKWIDKVIDA-DGYKNGFVNYKTG--EIEWKL 610  
 QY 285 ILAGND-----T-EDIKNFN-G---I-DGI-----A--PNAQIFSY---KMTSD 318  
 DB 611 IF--NDSSKLSIKSPTE-D-SLNSGQTFIDSLIHKIDISATP--QVGLIPEPNY-D 662  
 QY 319 AGSGFA-----GDETW---F-----HAIE---DS---I-KH---N--V--D---VVS--- 345

DB 663 V-T-FTKKENGNEQMLITFKKPLIPVEVYTKTPVGTITPLYNKRAVISDGEVLADYE 720  
 QY 346 --V-----SSGETG-TGL-VEEK--YWOAIRALRKAG--IP-MVVA-----TG-NVATSA--- 385  
 DB 721 AEVIDDNANKRVNKSQVGDNDIME--IVANQ--SGSTVSNAVTYDTIGTGKLTSLTKV 778  
 QY 386 --S-S-SWDLVA--NNHAKMT--D--TG-----N-VTRT--AAHEDAI 416  
 DB 779 YKSQTSVTKMKLQSN--MPSISPEYDLKTGVDEESNLEFQVFKNEINGSYVAKYQAI 837  
 QY 417 AVAS-AKN--Q---TVEP--DKVNI--GGESEFKRNI-----GAFD-----KS-KITTN 455  
 DB 838 TLTSDETTLTQIGNSVFTGPD--NITGGEKTKNIEVKITTYG---DGTGGEFKIILN 892  
 QY 456 E-DGTRK-PS-KLK---F-VY-----I-----GK-G--Q--DQDLI--GLD-----G 485  
 DB 893 KVD--KADPSIPLEGATFDLYANDEKVDOTTTDNGVIEFD-DLYVG-DYTLKVSAPEG 948  
 QY 486 -----K-IAVW---D---R-----I-----YTK-D-----LKNA-FKKAM- 506  
 DB 949 YTLPTASTENIQVLEBQDERKVQVMEKMPKENGVEHLVTKDKATGATLAGAEF--SLY 1006  
 QY 507 DK-GA--RAIMVY--N---TV-N-----YNRDNWTLBPANGYADEGTS--QVPSI-SG 550  
 DB 1007 DKSGAEIQLNGILTDENGELTIHNDLGSYYLKE--TRAPE-GYKLSE--KTWE-FSVESG 1060  
 QY 551 -D-----D-G-----V-----KL-----NMINPD--K--KT-----EVK 569  
 DB 1061 QVDAIEIOAENENKOLGAVLTKVDSFTNAKLSGAKFNLLN--DSGEVYQTNLVSDENGEI- 1118  
 QY 570 R-RNKE--D--FKDKLEQ---Y-----YF--I-----DM--ESFSNN--KPNV-- 598  
 DB 1119 RVQULBEGODYAFQ--ETLAPNTYDATTWPTIYVAGTSATMTVAE--N-NKTKSPDVT 1174  
 QY 599 G-----D---BK-B--I--DFKAPD-----TDK--E-----L-----Y--KEDI 621  
 DB 1175 GEVILVQDSATGSETLGAVALDMLTA--DGAIVASNLTTDANGELTYTNLAPGKYSFKS-T 1232  
 QY 622 IVPAG---STS-W-----GP-R---IDL-----LKPVSAP--G--KNIX-STLN-V--- 655  
 DB 1233 KAPGEYELATDVWFFETIAPNQPEKITTTAENTKLAP-I--PDAGSVKIIKODSENGVRLA 1289  
 QY 656 -----I--NGKS--T-Y-----G-Y 664  
 DB 1290 GAERSLAENGETLQTNLKTDENGEL 1315  
 RESULT 11  
 AAM09255  
 ID AAM09255 standard; Protein; 1938 AA.  
 XX  
 AC AAM09255;  
 XX  
 DT 27-MAR-1997 (first entry)  
 XX  
 DE Bacillus alkaline amylolupulinase.  
 XX  
 KW Amylolupulinase; alpha-amylase; pullulanase; detergent additive.  
 XX  
 OS Bacillus sp. KSM-Ap1378 (FERM BP-3048).  
 XX  
 FH Key location/Qualifiers  
 FT Peptide 1..32  
 FT Protein /label= Sig\_peptide  
 FT /label= 1938  
 FT /label= Mat protein  
 FT /note= "mature amylolupulinase"  
 FT Domain 33..833  
 FT /label= Alpha-amylase  
 FT /note= "N-terminal alpha-amylase"  
 FT Active-site 462..645  
 FT /note= "amy1ase active centre"

Query Match	34.3%	Score 1509;	DB 17;	Length 1938;
Beat Local Similarity	25.1%	Pred. No. 6.1e-16;		
Matches	292;	Conservative 166;	Mismatches 128;	Indels 578; Gaps 240

1 Y-PV-VLA--D-TSSS-E--D-AL-----ISDKE-KVAENKE-K-HENIHA 36

Db	25	YLPAAPAAETGDKRIEFSYERPDGNYEGMNLWMGTVGKDQIDFTBEKFGKAYADI--A	82
Qy	37	METSDQFEKKTJA-VI-K-----EKV-----VSNKPVDN-N-----TSNEAK--IKER	77
Db	83	--VS-D-NADKVGPIRKGDWBEKDPODDBRITINK-IDNTTKAVYISQOE-KEQOI-PD	135
Qy	78	NSNK--SGGDYDSFV-NKNTEN-PKKE-DKV-V-Y-I-AEFKDKE-SGEKAIXE----	122
Db	136	GSPVAVADGN-ADFFFRDK--ELVAAEGMDKYEKVELSILDE-K-YENNGEPE-KERTFY	189
Qy	123	-LSSIKNT-KLYTY-----D-R--I-----FN---G-SAIE-TT-----P--D-N	150
Db	190	TLSDLP-TGEHEYYLYLVTDQTEBEVDPYNTVDGRSVVEVYTSOVVSASFIPAKVYN	248
Qy	151	LDK-1K-OIEISSVERAKQVPM-MNHAKXI-GVEEA-ID-Y-L-----K-SINAPF	196
Db	249	QNAVYKVDIE--SEFE-TKIREMSINIS--EIGKEKATIDPALNELTVAKQGVTA--	300
Qy	197	G-KNPDGKGVISNID-TGTDYRHK-A-----MR-I-----D-D-DAKA--SM--RFX	234
Db	301	GVKNLP-----ITAIDEFGN--RHEGSALEVOQARTINGEKADPFMDQSVVYFMLTDHFF	353
Qy	235	KEDLKCTD-KYMYISDKIPKAFNV-----YNGG-K-ITVEKTD-D-G-----	271
Db	354	--D-G-DSSN--ND--PHGIGDYTSKSGTYOGDGFKGIT-ORLUDYDELGINTIWISP	402
Qy	272	-----R-----D-YF-----D-----PH-G-M-----H-I-A--GI-LAGN	289
Db	403	VVDNIKPVNRHSEGBDPTPYAYHGYMADNGELNPHFGSMADFHEMIDAAHERGKIKN-V	461
Qy	290	DTEODIKN--FNGI--D-GIA-PN-----AOIFSYKMYSDAGSG-----PAG--DET	328
Db	462	DV---VLNHTGY-GLKPEDSSVANFPDEDDRAR-FD-CMLRDGSSGVRBELAGLPD--	513
Qy	329	MFAHIE-DSIKANV-----D-V-VS-VSSG---P-TGT-GLVGEK-YQOAIR-ALRK	369
Db	514	-FLT-ENPD-VREQVQWQOTWIEKRSRFAKNNTIIDYFVDVIVKHY-EDTTMAAFKALTLK	569
Qy	370	AGIP-WWV-AT-G--N-----YATS--ASS-----SSW--DLVANHNL-----KM--	400
Db	570	A-MPHKILIGEMKANVNDLGLYNLSGMSDLDLDFKNYARDP-ANQOLDAVOOKLEAR	627
Qy	401	-TDTGN-VT-R-TAAH-ED-----AI-----AV-AS-----AKNQVY--EP-----	428
Db	628	NSKLNNTATLQGLFSGHEDRFYEVBEDDLKRYAASLQLTALGQPIYYIGBELGLGK	687
Qy	429	-D-KV-----N-----I-GGSEFK-YRNIGAF-----F--DKSKITTNEDGTAKDSK	464
Db	688	NDYPYTTRQMPMDVDGNEILEHYQTLAFRDNPTPAFKGRKKYA-----GSDSEGY	743
Qy	465	LKF-----VYIG--KGODDLIGLDLUG--KIANMDITYT-KDL-KN-----ARK-	503
Db	744	LLFRTYGENSVYGLNTEAALAKD-VTLNF-GSSEAVTDR-YSGOEYQANEEGQVTESI	800
Qy	504	KAM-DKGARAIMV-V-N-----T-----V-N-N-Y--YNR-DN-----W-----	527
Db	801	PAMED-GG-TYLLVENKANVPVEEPEPPEIEBNTLARIHQRTDINSYENUGLMDQV	858
Qy	528	--TE-----LP-AMG-----YEA-D-EGTK-SQ-V-FSI--S-GD-DG---VKL--	556
Db	859	AAPSEBNMSGTPOAGNVTDYGAIVVDVLAEDQNIIGFLVLTNTNGCKDGGDKAVELFS	918
Qy	557	-----MM-----I--NP-D-KKT-EV--KXNNKEDPRD-KLEQY-YPIDMSFPMN	594
Db	919	PDLNEIWIKQSGDEFLYEPVDLPAINTVRIHYERN-ADYEGWGL--NMWR-DVES-DS	973
Qy	595	K-PN-----VG-----D-E-KEID-----PKFA-P-----D-T-D-K--ELY-	617
Db	974	GMPGACADAAGIATGKGAAYDIKLKE-DANKIGFLPVAKQSGGQGDMPFMDLKQYNQJFV	1033
Qy	618	K-ED-I-----IYP-A--G-----STSGPR--IDL-L--LK-PDVSAPGK	647
Db	1033	KEGDKQVYTNPGYVPLALVSGEVLSDKLSTLPT-TRREGDIDELKQELIKQVD--G-	1088



XX ABP30377;  
AC 02-JUL-2002 (first entry)  
DT XX  
DE Streptococcus polypeptide SEQ ID NO 9930.  
KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
KM antiinflammatory; infection; vaccine; meningitis; gene therapy.  
XX  
OS Streptococcus agalactiae.  
XX  
PN W0200234771-A2.  
XX 02-MAY-2002.  
XX  
XX 29-OCT-2001; 2001WO-GB04789.  
XX  
XX 27-OCT-2000; 2000GB-0026333.  
XX 24-NOV-2000; 2000GB-0028727.  
XX 07-MAR-2001; 2001GB-0005640.  
XX  
XX (CHIR-) CHIRON SPA.  
XX (GENO-) INST GENOMIC RES.  
XX  
XX Telford J, Maignani V, Margarit Ros YI, Grandi G, Fraser C,  
XX Tettein H;  
XX  
XX WPI: 2002-352536/38.  
XX N-PSDB; ABN71008.  
XX  
XX New Streptococcus protein for the treatment or prevention of infection  
XX or disease caused by Streptococcus bacteria, such as meningitis, and  
XX for detecting a compound that binds to the protein -  
XX  
XX Claim 1; Page 4122; 4525pp; English.  
XX  
XX The invention relates to a protein (ABP25413-ABP30895) from group B  
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
XX (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
XX the specification. The proteins have antibacterial and antiinflammatory  
XX activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and  
XX antibodies that bind (1) are used in the manufacture of medicaments for  
XX the treatment or prevention of infection or disease caused by  
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
XX Nucleic acids encoding (1) are used to detect Streptococcus in a  
XX biological sample. (1) is used to determine whether a compound binds to  
XX (1). A composition comprising (1) or a nucleic acid encoding (1), may be  
XX used as a vaccine or diagnostic composition. The disease caused by  
XX Streptococcus that is prevented or treated may be meningitis. Nucleic  
XX acid encoding (1) may be used to recombinantly produce (1) and may be  
XX used in gene therapy. Antibodies to (1) are used for affinity  
XX chromatography, immunoassays, and distinguishing/identifying  
XX Streptococcus proteins.  
XX  
SQ Sequence 2274 AA;  
Query March 34.2%; Score 1506; DB 23; Length 2274;  
Best Local Similarity 25.6%; Pred. No. 1.5e-15;  
Matches 280; Conservative 169; Mismatches 150; Indels 493; Gaps 222;  
QY 1 YP--VV-L-AD--TSSSEDA-LN---I---SDKEKVAE-NK-----EK-HENIH 34  
DB 458 YPLGSLVSYKQDEFWMSVS-DARLNGLRIRIELVNDPSDITEQNPVLVYRWEEVSQALH 516  
QY 35 --SAMETSQ-DFKE-KK-T--AVIKEEV--VS--KNPVID--N--N-TSNEEA--KIK 75  
DB 517 QPKA-EP-QTELEADQELNLFSLEREPVQSIGLE-PD-DSKNGHNDTDLJETDQIP 572  
QY 76 EENENKS--QG--D-Y-----TDSF-----V--N-----KN--TE--N--P-KK 100  
DB 573 EEEVEETIPEIPTVDFPEEDLTD-FYPKTARDKVEETNIVAIRLVKQLEVHRNASPSEQ 631

QY 101 E--DKVY-Y--IA-E-EKD-----KESGE-KA-I--KELSSLKNTK--VLVT----- 134  
DB 632 ELAKYVGMGGLANFEFDYDNPFKSKEREELKSLVTDKXYSMDKQSSLTAYTDPSSLIRQ 691  
QY 135 -YD---R-INGSAIETTPD-----N-----LDK-IKQ-IE--GI--SSVEEA-QK-VOPMM 173  
DB 692 MMDLEREDGTGGKI-LDPSMGTFNFPAMPKHLREKSELYGVELDTITGAIAKHLHP-- 748  
QY 174 N-HARKRI-GVEEAI--D--Y-LKSI-NAPF-----GKU-----F----- 200  
DB 749 NSHI--EIKGF-ETVAANDNSFDL-VISNVFANIRIADRRYDRPYMIDYFVAKSIDDL 804  
QY 201 -DGRGV-ISNI-DTGT-D-----Y--RH--K--AMRIDDA-KA-----S-----M-R 232  
DB 805 HDG-GQVAL--ISSTGMDKRTENILGDIRTTEPLCGVRLPDSAPALAGTSTVTMLMF 861  
QY 233 F-K--K---EDL--KGT---DKN--YWLSDKI PHAF-----N-----Y--YNGCKI 263  
DB 862 FOKHLDKGVVADDLAFSGSIRYDKSRIML--NPF--FDGBYNSQVLGTVEVNFNGITL 917  
QY 264 TVEKYDDGRDYPD--PHGM-HI-AG--I-----LAGN-D--TBO-D--I-----KN----- 297  
DB 918 SVKGTSD--DLIASVETALNHVKAPEIDRNEVYINPDVLTKOVNDTSIPAEMRENLGQY 975  
QY 298 -FN--G--I--D--GI---APNAQIYSYKQVSD-AGSGF-AGDETWFH--AIED---S 336  
DB 976 SFGYQSTVYRRBNKGRVGTKEEII-SY--YDDEG-NFKAMD-TK-HGQKQ-DKFNPA 1028  
QY 337 IK--HN--VDV-VS--G-FTGTGLVEEK--YMOA-IRALRK-AGIP-MV-VATGN- 380  
DB 1029 LEVTDNTALVDVYDDAKKRGQF--KGYV-KKTVFYFAPV-SYKAVARIMGVDIR--NA 1082  
QY 381 YA-TSASSSMDLV--ANNHL--KM--T-DT-----G--N--YTR-----T-AAH 412  
DB 1083 YQEVIALQRYDYDIDKETFNHLGKLNRTYDSFVKHGYLNSAVNRNLPDSDDKXSLASL 1142  
QY 413 ED-AI-----AV--AS-A-KNOTV-----B-----FDKVN--I--G-GESEFY-RNIG 444  
DB 1143 EDSLSOPSGKSVIYTSLSLAPEKALVREKEVKVYHTALDLNLSLADGKRVDFAYMMSI- 1201  
QY 445 AF-PDKSKITTED-G--TKAPSK-L-KFYVYGQDDLI-GLDLRKXIAMDRIYT 495  
DB 1202 -YGV-ESQMTLIBELGDLIMPDEKYLNGELTYV--SR-ODPLSG-DVVTKLEVVD-LFV 1254  
QY 496 K-DLKNA-EKKANDKG-----ARAIMVNTYNY--NNDNTELP-----A-----M 533  
DB 1255 KOD--NODFWMSHYAGLLEAIKPAR-ITLAD-IDYRIGSR--W--IPLAVYGRFAOETFM 1306  
QY 534 G--YE-AD-EGTSQVPSISGDDGV-----KL-WNMIN-PDKTTEVK--RNKED----- 575  
DB 1307 GKAYELSDQE-VAT-VLEVPIDGVITYGSKPATYTSNATDRSLGCVASR---YDSRKI 1361  
QY 576 FKDKL--EQYYP-IDMESF-NSNKNPVGD-E-----KE--I-D-FK-F-A--P----- 610  
DB 1362 FENLINSNQ--PITTKOVNKGDKKKNVTDVEKTTVLRKETHLQELFGVAKPBEVQOM 1419  
QY 611 --DTDKELYKEDIIVPAGSTSW-G-PR-IDLLKLPDVS-AP-GKN-----I---K-STLN- 654  
DB 1420 IETPTNLYNR--TV---SKSYDGSHLTD-GLAONISLRPHOKNAIQRIVBERALLAH 1473  
QY 655 -VINGKS-T-YG 663  
DB 1474 EVGSGKTLTMLG 1485  
RESULT 14  
ABP28340  
ID ABP28340 standard; Protein; 2278 AA.  
XX  
AC ABP28340;  
XX  
DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 5856.  
 XX  
 XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;  
 KM Streptococcus; Streptococcus pyogenes; antibacterial;  
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;  
 KM anti-inflammatory; infection; vaccine; meningitis; gene therapy.  
 OS Streptococcus agalactiae.  
 XX  
 XX WO200234771-A2.  
 XX  
 XX 02-MAY-2002.  
 XX  
 XX 29-OCT-2001; 2001WO-GB04789.  
 XX  
 XX 27-OCT-2000; 2000GB-0026333.  
 XX  
 XX 24-NOV-2000; 2000GB-0028727.  
 XX  
 XX 07-MAR-2001; 2001GB-0005640.  
 XX  
 XX (CHIR-) CHIRON SPA.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 XX Telford J, Maignani V, Margalit Ros YI, Grandi G, Fraser C;  
 PI Tectel H;  
 XX  
 XX WPI; 2002-352536/38.  
 DR N-PSDB; ABN68971.  
 XX  
 XX New Streptococcus protein for the treatment or prevention of infection  
 PT or disease caused by Streptococcus bacteria, such as meningitis, and  
 PT for detecting a compound that binds to the protein -  
 XX  
 XX Claim 1; Page 3748; 4525dp; English.  
 PS  
 XX The invention relates to a protein (ABP25413-ABP30895) from group B  
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
 CC the specification. The proteins have antibacterial and anti-inflammatory  
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
 CC antibodies that bind (I) are used in the manufacture of medicaments for  
 CC the treatment or prevention of infection or disease caused by  
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
 CC biological sample. (I) is used to determine whether a compound binds to  
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
 CC used as a vaccine or diagnostic composition. The disease caused by  
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
 CC acid encoding (I) may be used to recombinantly produce (I) and may be  
 CC used in gene therapy. Antibodies to (I) are used for affinity  
 CC chromatography, immunoassays, and distinguishing/identifying  
 CC Streptococcus proteins.  
 CC  
 XX  
 XX Sequence 2278 AA;  
 SQ  
 Query Match 34.2%; Score 1506; DB 23; Length 2278;  
 Best Local Similarity 25.6%; Pred. No. 1.5e-15;  
 Matches 280; Conservative 169; Mismatches 150; Indels 493; Gaps 222;  
 QY 1 YP---VV-L-AD---TSSSEDA-IN---I---SDKEKVAE-NK-----EK-HENIH 34  
 DB 462 YPLGSLVSYKQDPEFMSVS-DARLNGLRIRLAVDFSDIIRQNPVLVYRTVEEVSQALH 520  
 QY 35 --SMEETSO-DFKE--KK-T-AVIXKEV--VS--KNPVID--N--N-TSNEEA--KIX 75  
 DB 521 QPKA-EP-QTELEBADOELNLFSLFEEBPVOSIGLLE-PD-DSENGNDTDLJEETDNOIP 576  
 QY 76 EENSNSK--QG---D-Y-----TDSF-----V--N-----KN--TE--N--P--KK 100  
 DB 577 EEEVETIPEIPIVNDYFPEPDLTD-FYPRKADKQETNIVAIKLVKNLEVEHRNAPSQ 635  
 QY 101 E-DKVV-Y--IA-E-FPD-----KESGE-KA-I--KELSLKNTK--VLYT----- 134  
 DB 636 ELAKYVGMGLANFEFDYDNPFPKSKEREELKSLVTDKXEYSDMKOSLTAAYTDSLIRQ 695

QY 135 -YD---R-IFNGSAIETTPD-----N-----LDK-IKQ-IE--GI--SSVERA-QK-VOPWM 173  
 DB 686 MMDKLERDGFITGCKI-LDPSMGTONFPAAMPKRLHREKSELVGVELDTITGALAKLHP-- 752  
 QY 174 N-HARKEI-GVEBAL--D-Y-LKSI-NAPF-----GKN-----F----- 200  
 DB 753 NSHI--EIKGF-ETVAFNDNSFDL-VISNVFANIRIADNRYPMTYMDYFVKXSLDLL 808  
 QY 201 -DGRGMV-ISMNI-DTGT-D-----Y---RH--K---AMRIDDDA-KA-----S-----M-R 232  
 DB 809 HDG-GQVAH--ISSGTMDKRTENILQDIRETTELGGVRLPDSAFKALAGTSVTTDMIF 865  
 QY 233 F-K--K-----EDL--KGT--DKN--YMLSDKIPHAF-----N-----YNGSKI 263  
 DB 866 FOKHLDKGVAVDDLAFLSGSIRYDKDSRIWL--NRY-FDGEVNSQVLGTYEVRNFGTL 921  
 QY 264 TVEKYDDGRDYFD--PHGM-HI-AG-I-----LAGN-D--TEQ--D--I-----KN----- 297  
 DB 922 SVKGTSD--DLIASVETALNHVKAPREIDRNEVIINPVLTQVNDTSIPAMRENTLQY 979  
 QY 298 -FN--G--I---D--GI---APNQIFSYKMSD-AGSGF-AGDETMFH---AIED---S 336  
 DB 980 SFGYQSTVYRDNKGIKRVGKTBEI-SY--YVDEEG-NFKAMD-TK-HSQKQI-DRENA 1032  
 QY 337 IK--HN--VDV-VS--VSS-G-FITGGLVGEK--YMOA-IRALRK-AGIP-MV-VATGN- 380  
 DB 1033 LEVTDNTALDYVYDDDAKRGQF--KGYI-KKTVEYEAFL-SYKEVARIKGVNDR--VA 1086  
 QY 381 YA-TSASSSMDLV--ANNHL--KM--T-DT-----G--N--VTR-----T--AAH 412  
 DB 1087 YQEVIAIQRYVDYDDETENHLGKLNRTYDSFVKHYGYLNSAVNNHLPDSDDKYSLSLAL 1146  
 QY 413 ED-AI-----AV---AS-A-KNQTV-----E-----FDKVN--I--G-GESPKY-RNIG 444  
 DB 1147 EDESLDPSGKSVIYTKSLAFKALVRPREVKKCHTALDALMSLADGKGVDFATYMSI- 1205  
 QY 445 AF-FDKSKITTNE-D--G--TRAPSK-L--KFVYIGKGDODLI-GLDLRGKIAVMDRIYT 495  
 DB 1206 -YQV-ESQMTLIEELGDLIMDPBKYLNGELTYV--SR-QDPLSG-DVYTKLEVVVD-LFV 1258  
 QY 496 K-DLKA-FKAMDKG-----APALMVNTVNY--NRDWTLP-----A---M 533  
 DB 1259 KQD--NQPNNSHYAGLEIAIKPAR-ITLAD-IDRIGSR--W--IPLAVGKFAQETPM 1310  
 QY 534 G--VE-AD-EGTKSVFSGDDGV-----KL--NMNIN--PKKTEVK--RNKED----- 575  
 DB 1311 GAYELSDQE-VAT-VLEVSPIDGVITYQSKFATYVSNATDRSLGVPSR---YDSGRKI 1365  
 QY 576 FKDKL--EQYYP-IDMESF-NSNKEPVGD-E-----KE--I-D-FK-F-A--P----- 610  
 DB 1366 FENLANSNQ--PTIKQYVEGDKKKNVDVEKTTVLAKEHTLQGLFQGFVAKYEVQOM 1423  
 QY 611 --DTDKELYKEDIIVPAGSTSW-G-PR-IDLLIKPDVG-AP-GKN---I---K-STLN- 654  
 DB 1424 IEDTYRNLNLR--TV---SKYDGSHTLTD-GLAONISLRPHQKNAIGRIVEKDALLH 1477  
 QY 655 -VINGKS-T-YG 663  
 DB 1478 EVSGSKTILMLG 1489  
 RESULT 15  
 AAB10667  
 ID AAB10667 standard; Protein; 2057 AA.  
 AAB10667;  
 AC AAB10667;  
 XX  
 XX 19-JAN-2001 (first entry)  
 DT  
 XX  
 DE L. mesenteroides alterman sucrose protein.  
 XX  
 XX Alternan sucrose; glucosyltransferase; fructose; cosmetic; foodstuff;

KM BYRUP.  
 XX Leuconostoc mesenteroides.  
 OS DB19905069-A1.  
 XX  
 XX  
 XX 10-AUG-2000.  
 PD  
 XX 08-FEB-1999; 99DE-1005069.  
 XX 08-FEB-1999; 99DE-1005069.  
 PR 08-FEB-1999; 99DE-1005069.  
 XX  
 PA (PLAN-) PLANTTEC BIOTECHNOLOGIE GMBH.  
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PI Koesmann J, Weissh T, Quanz M, Knuth K;  
 XX WPI: 2000-550294/51.  
 DR N-PSDB; AAA97904.  
 XX  
 XX New nucleic acid encoding recombinant Leuconostoc mesenteroides  
 PT alternan sucrose protein and methods of alternan and fructose  
 PT production -  
 XX  
 PS Claim 1a; Page 30-36; 64pp; German.  
 XX  
 CC This invention describes a novel nucleic acid molecule (I) encoding an  
 CC alternan sucrose (E.C. 2.4.1.140 - an enzyme, that belongs to the  
 CC glucosyltransferase group) The recombinant, purified alternan sucrose  
 CC gene is useful for the fermentative production of alternan (a  
 CC carbohydrate) and/or fructose by secreting the enzyme into a  
 CC saccharose-containing culture medium. Alternatively, the enzyme is  
 CC contacted with a saccharose-containing solution. The alternan and/or  
 CC fructose is then isolated from the medium. Cosmetic products or  
 CC foodstuffs containing alternan can be produced. Recombinant production of  
 CC alternan sucrose is advantageous as it provides a cost effective means of  
 CC producing fructose for high fructose containing syrups, production of  
 CC which previously has been achieved by costly production from maize  
 CC starch. This sequence represents the Leuconostoc mesenteroides alternan  
 CC sucrose protein which is described in the method of the invention.  
 CC  
 XX  
 SO Sequence 2057 AA;  
 Query Match 34.2%; Score 1502; DB 21; Length 2057;  
 Best Local Similarity 24.3%; Pred. No. 1.1e-15;  
 Matches 296; Conservative 150; Mismatches 155; Indels 617; Gaps 240;  
 QY 1 YPV-VIADT--SSSE---DAL-N-ISDKEK--V--AE---NKEKHENHSAM--E 38  
 DB 63 VVPVSTNDSILKQGTDFWYDSGNRV-D-QKTQIILLTAELQKKNNEK--NL-SVISDD 117  
 QY 39 TS-ODPFEKKTAVIKKEKVVSKN-PVIDN---NTSNEAKI--KE-ENSKNGCD--YT 87  
 DB 118 TSKKD-DE--NISQKTKIA--NOQTVDYTKAGLTTSNPDITGAYENHN--GYVYI 168  
 QY 88 D-S-F-VN--KNTN-N-PKEDKVVY-I-ABFKDSEGEKAIKELSL--K--NTKVL- 132  
 DB 169 DASGKQVLTGQNTIDGNLQYFDN-GYGVKGSFRV-NG-KHIV-FDVTGKASANDVIVN 224  
 QY 133 -----Y-----TY--D--RI--F--NG-SAI--ETTPD-NL--DK--IKQIE-G 159  
 DB 225 GKAQGYDAQGNOLKKSIVADSSGQTYFDFGNGPLILQTI-DGNLQYFNGQGV-QIKGG 282  
 QY 160 ISSV--BRAQKVOPMM-NHARK--EI--G--VE--EAI-DVLKSNAPF-----GKN--F 200  
 DB 283 FQDVNNKRI-YFAPNTGN-AVANTETLNGKIQGKADANGOVK--NA-FSDVAG-NTPYF 336  
 QY 201 DGRG-WV--ISNIDTG-TDY--R-H--K--A-----NRIDD--A-YASM--RF-- 233  
 DB 337 DANGVMLTGLQTI-SGKTYLIDEGGHLRKQYVAGFPNNGFVFPADDTGAGKTAIEYQPDG 395  
 QY 234 ----KKE-----D--LK--G-----TD--KN--YVL-S--DK--I- 251

DB 396 LVQSQSNENTPHNAKSYDKSSFENVDDGYLTADTYWYRPTDILKNGDTWTASTETDMRELLM 455  
 QY 252 ---P---HAFNYN-----G-G-----K-----I-T-VE-----K---Y 268  
 DB 456 TWMPDKQTOA-NYLNPFSSKGLGTTTTTYTAATSOKTINDAAFYVQTALBQOISLKSTEW 514  
 QY 269 -D-D-----G-R-----D--Y-----F-----DPH-----GMH--IAG--- 284  
 DB 515 LRDAIDSFVKTQANWNNQTEDEAFDGLQMLQGGFLAVQDDSHRTPNTDSNNRL--GRP 573  
 QY 285 I-LAG-NDTEODIKN--F---NGIDGAPNAQI--FSYKMY-SDAGSGFAG--DETFFHA 332  
 DB 574 INIDGSKDT-TDKGSEFLLANDIDNSNPYQAEQMLMHLHYLNKFGS-ITGNNDNANFDG 631  
 QY 333 IE-DSIKENV--DVVSVSSG--F-T--GT-----G--L-VGBKVV-----Q-AIALR 368  
 DB 632 IRVDAYD-NVDADILKI-AGDYFKALYGTDKSDANANKHLSIBD--WNGDPOVANO--Q 686  
 QY 369 KAG---IPM--VYAT--GN-YAT-SAS-S-S-W---D--LVAN---NHLKPTD-----TG 404  
 DB 687 --GNAQLTMDYVTVTSQFNSL--THGANNRNNMYPIDTGYLLNGDLN-KKIVDKNRNSG 742  
 QY 405 N-VTR-----T-----A-AHE-DA--I--A-V--ASAKN-Q-TVEPDKNIGES 437  
 DB 743 TLVNRIRANSQDTVIVPVSFVRADYDAQPIPKAMIDHGIIRNMOPTFPFDOLAQGM- 801  
 QY 438 FKY--R-N-IGAF--F-D--KS---KITNED-----GT--KAP- 462  
 DB 802 FYKQGNBPG-FKKTINDVLPBAYVAMLTKQTVPRVYGYDMYLBEGQYMEKGTIYNPV 860  
 QY 463 -SK-LK-FV-YI-GKQD---D--ODLI-G-LDL---R-GKIAVW--DRIYT---KD 497  
 DB 861 ISALLKARIYVVG-QGYMATDSGKDLKNGERDILTSVAFK-GIMTSQTTQDMSQD 918  
 QY 498 LKNA-----FKKAMK-----G-A-----RAIMV-N-T-VNYN----- 523  
 DB 919 YKQGGIVYGNPNPDKLNNDKITTLHMGVAKHQLYRA-LVLSNDSGIDVYDSDKAPT 977  
 QY 524 -R--DNWTEL---P---AW--G---YEADGKTS-QVPSISGDDGVKLMNNINP----- 562  
 DB 978 LRTNDNG-DLIFPKTNTFVAKQDGTITNYEM-KG--SLNAL-IGGYLGV--W--V-PVGAS 1027  
 QY 563 ---DKK---TEVRANKED---F--KDKLE-QY-YPIDMESFNSS--KP-----NV- 598  
 DB 1028 DQDARTVATSSSSN--DSGVSHNAALDSNVIY---EGF-SNQAAPTSPESQSTNV 1080  
 QY 599 -GDE---KE-I-DFKAP-----DT-----DKEL--YK--E--DITV-PA-GS 627  
 DB 1081 IATKANLFXELGITSFELAPQYRSSGDTVYCGMSFLDSFLANGYAFTRDYLGFNRADGN 1140  
 QY 628 ---TSMGP---R--IDLK-----PD-VGA-FCK-----N-IKST- 652  
 DB 1141 PNPFTKGTQDLDNALBALHKNQMOALADWPQIYALPKKVVYVATRVDERGNOLKDTD 1200  
 QY 653 -LN---VINGKSTYV--Y 664  
 DB 1201 FVNLVAVANTKSS-GVDY 1217  
 Search completed: October 14, 2003, 16:30:18  
 Job time : 98 secs  
 <UL>  
 <LI><STRONG>pat</STRONG>  
 <UL>  
 <LI>US09590991  
 </UL>  
 <LI><STRONG>dir</STRONG>  
 <UL>  
 <LI>runat\_14102003\_154349\_14329  
 </UL>  
 <LI><STRONG>printer</STRONG>  
 <UL>  
 <LI>0



```
</UL>
<LI><STRONG>html_save</STRONG>
<UL>
<LI>0
</UL>
<LI><STRONG>runat_14102003_154349_14329</STRONG>
<UL>
<LI>US-09-590-991-6
</UL>
<LI><STRONG>save</STRONG>
<UL>
<LI>Save AB.
</UL>
</UL>
IN=/cgn2_1/USPTO_gpool/US09590991/runat_14102003_154349_14329/output, F=/cgn2_1/USPTO_gf
```

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using BW model

Run on: October 14, 2003, 16:26:41 ; Search time 48 Seconds  
(without alignments)  
2195.718 Million cell updates/sec

Title: US-09-590-991-6  
Perfect score: 4398  
Sequence: 1 YPVVLADTSSSEDALNISDR.....PGKNIKSTLVINGKSTVGY 664

Scoring table: BLOSUM30  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.\*  
23: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.\*  
24: /SIDSL1/gcgdata/geneeq/geneeqp-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4398	100.0	664	22	AA198342
2	4398	100.0	2120	21	AA1981710
3	4379	99.6	2140	24	ABU01020
4	4171	94.8	1007	19	AA161246
5	4171	94.8	1007	23	ABP54664
6	974	22.1	192	13	AA186159
7	899.5	20.5	1962	12	AA105559
8	899.5	20.2	1962	12	AA10557
9	899.5	20.2	1968	12	AA10941

10	888.5	20.2	1962	12	AA10558	Mutant protease (A
11	887.5	20.2	1974	12	AA10940	Mutant protease (d
12	885.5	20.1	1962	12	AA10560	Mutant protease (K
13	885.5	20.1	1962	12	AA10561	Mutant protease (N
14	885.5	20.1	1962	12	AA10563	Mutant protease (K
15	884	20.1	1959	12	AA10562	Mutant protease (d
16	866.5	19.7	1560	10	AA194145	S. cremoris protei
17	817.5	18.6	1946	19	AA147273	Lactobacillus bulg
18	804.5	18.3	1570	23	ABP29894	Streptococcus poly
19	804.5	18.3	1570	23	ABP30531	Streptococcus poly
20	804.5	18.3	1590	23	ABP28560	Streptococcus poly
21	802.5	18.2	1550	23	ABP30895	Streptococcus poly
22	793	18.0	1579	24	ABP71293	Group B Streptococ
23	786	17.9	885	23	ABP27332	Streptococcus poly
24	786	17.9	885	23	ABP29796	Streptococcus poly
25	773	17.6	1647	23	ABP25822	Streptococcus poly
26	768	17.5	1239	23	ABP25822	Streptococcus poly
27	767	17.4	1233	23	ABP25822	Streptococcus poly
28	767	17.4	1233	23	ABP30203	Streptococcus poly
29	707	16.1	806	13	AA127481	RP-III residual pr
30	701	15.9	690	11	AA104904	Residual protease-
31	691	15.7	690	18	AA16337	Bacillus subtilis
32	691	15.7	690	18	AA101787	Residual protease
33	686	15.6	815	23	ABP91806	Herbicideally activ
34	669.5	15.2	1167	18	AA12470	SCP12 peptidase (
35	668.5	15.2	1167	18	AA101264	SCP12 peptidase (
36	667.5	15.2	1164	21	AA12469	SCP12 peptidase (
37	665.5	15.1	1164	21	AA101263	SCP12 peptidase (
38	661.5	15.0	1150	21	AA101265	SCP12 peptidase (w
39	660.5	15.0	1150	18	AA12471	SCP12 peptidase (w
40	657.5	14.9	1181	21	AA101266	SCP12 peptidase (w
41	657.5	14.9	1181	21	ABP25823	Streptococcus poly
42	652.5	14.8	1114	23	ABP29883	Streptococcus poly
43	645.5	14.7	1090	23	ABP298458	Streptococcus poly
44	641	14.6	903	17	AA187007	Hyperthermostable
45	641	14.6	922	19	AA17372	Trimmed enzyme pro

## ALIGNMENTS

RESULT 1	AA198342	standard; Protein; 664 AA.
XX	AA198342:	
AC	20-APR-2001	(first entry)
XX		
DT		
XX		
DE	S. pneumoniae Sp128 polypeptide.	
XX		
KW	Immunogenic; Sp128; Sp130; pneumococcal; otitis media; nasopharyngeal;	
KW	bronchial; lung; blood; infection; immune response; immunotherapy;	
KW	antibacterial; auditory; vaccine.	
XX		
OS	Streptococcus pneumoniae.	
XX		
PN	W0200076540-A2.	
XX		
PD	21-DEC-2000.	
XX		
PF	09-JUN-2000; 2000MO-US15925.	
XX		
PK	10-JUN-1999; 99US-0138453.	
XX		
PA	(MEDI-) MED IMMUNE INC.	
XX		
PI	Adamou JE, Choi GH;	
XX		
DR	WPI, 2001-112197/12.	
DR	N-PSDB; AAC84741.	
XX		
PT	New vaccines comprising Sp128 or Sp130 polypeptides, for treating and	

PT preventing pneumococcal infections, particularly infections caused by  
 PT Streptococcus, e.g. otitis media, nasopharyngeal, bronchial, lung or  
 PT blood infections

XX Claim 4; Page 47-50; 54pp; English.

XX The invention relates to novel immunogenic polypeptides, Sp128 and Sp130  
 CC from *S. pneumoniae*. Vaccines comprising the polypeptides are useful for  
 CC the treatment and prevention of pneumococcal infections, particularly  
 CC infections caused by *Streptococcus*, such as otitis media, nasopharyngeal,  
 CC bronchial, lung or blood infections. The antigens are used as immunogenic  
 CC agents to stimulate an immune response. The antisera and antibodies may  
 CC also be used in diagnosing and treating pneumococcal infections.  
 CC Recombinant polypeptides serve as a mechanism for stimulating production  
 CC of antibodies for use in passive immunotherapy, diagnostic reagents, and  
 CC as reagents in other processes such as affinity chromatography. The  
 CC present sequence represents the *S. pneumoniae* Sp128 polypeptide.

XX Sequence 664 AA;

Query Match 100.0%; Score 4398; DB 22; Length 664;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-60;  
 Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNLSDEKVAENKEKHENHSAMETSODFKEKTAIVIKEKEVSKN 60  
 DB 1 YPVVLADTSSSEDALNLSDEKVAENKEKHENHSAMETSODFKEKTAIVIKEKEVSKN 60  
 QY 61 PVIDNNTSNEBAKIKEENSNKSGDYTDSFVNKNTENPKKEDKVYVIAEFKDSGEKAI 120  
 DB 61 PVIDNNTSNEBAKIKEENSNKSGDYTDSFVNKNTENPKKEDKVYVIAEFKDSGEKAI 120  
 QY 121 KELSLKNTKVLTYTDRIFNLSAIEETPDNDLKIQIIGISSVERAQVQPMNHARKKI 180  
 DB 121 KELSLKNTKVLTYTDRIFNLSAIEETPDNDLKIQIIGISSVERAQVQPMNHARKKI 180  
 QY 181 GVEBAIDVLKSIINAPFGKPFDRGMVINSIDTGTDRYRKAMRIDDDAKASRFKEDLKG 240  
 DB 181 GVEBAIDVLKSIINAPFGKPFDRGMVINSIDTGTDRYRKAMRIDDDAKASRFKEDLKG 240  
 QY 241 TDKNYWSLSDKI PHAFNYNGSKITVEKYDDGRDYDPHGMHAGILAGNDTBODIKFNNG 300  
 DB 241 TDKNYWSLSDKI PHAFNYNGSKITVEKYDDGRDYDPHGMHAGILAGNDTBODIKFNNG 300  
 QY 301 IDGIPNAQIFSYKMYSDAGSGFADDETFPHAIEDSIRKNDVVSVSSGFTGTGLVGEKY 360  
 DB 301 IDGIPNAQIFSYKMYSDAGSGFADDETFPHAIEDSIRKNDVVSVSSGFTGTGLVGEKY 360  
 QY 361 WQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDTGNVTRTAHEDAIIVAS 420  
 DB 361 WQAIRALKRAGIPMVVATGNVATSSASSSSWDLVANNHLMKMTDTGNVTRTAHEDAIIVAS 420  
 QY 421 AKNQVVEEDKNIIGESFRYRNIGAFPDKSKITTNEDGTAKPSKLFYVIGKQDQDLIG 480  
 DB 421 AKNQVVEEDKNIIGESFRYRNIGAFPDKSKITTNEDGTAKPSKLFYVIGKQDQDLIG 480  
 QY 481 LDLRGKIVMBRIYTKDLKNAFFKAMDGARAIVMNVVNYNRRNMTLELPMAGEADEG 540  
 DB 481 LDLRGKIVMBRIYTKDLKNAFFKAMDGARAIVMNVVNYNRRNMTLELPMAGEADEG 540  
 QY 541 TKSQVFSISGDGVLAMNINPDKTEVYRNKKEDEPKLEQYPIIDMESFSNKNPNVGD 600  
 DB 541 TKSQVFSISGDGVLAMNINPDKTEVYRNKKEDEPKLEQYPIIDMESFSNKNPNVGD 600  
 QY 601 EKEIDFKAPPTDKELYEDIIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLVNINGS 660  
 DB 601 EKEIDFKAPPTDKELYEDIIIVPAGSTSWGPRIDLLKPDVSAFGKNIKSTLVNINGS 660  
 QY 661 TYGY 664  
 DB 661 TYGY 664

RESULT 2  
 ID AAY81710 standard; Protein: 2120 AA.

XX AAY81710;

DT 02-JUN-2000 (first entry)

XX Streptococcus pneumoniae protein sequence ID3.

XX Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;  
 KW bacterial pneumonia; asplenia; heart disease; lung disease; alcoholism;  
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;  
 KW pneumococcal septicemia; sinusitis; meningitis; therapy.

XX Streptococcus pneumoniae.

XX W0200006738-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99MO-GB02452.

XX 27-JUL-1998; 98GB-0016336.

XX 19-MAR-1999; 99US-0125329.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Le Page RWF, Wells JM, Hambley SB, Hansbro PM;

XX WPI; 2000-195301/17.

XX N-PSDB; AA291806.

PT Streptococcal proteins and polynucleotides useful for diagnosis,

PT treatment and prophylaxis of bacterial infections

XX Claim 2; Page 41-42; 76pp; English.

XX This sequence represents a *Streptococcus pneumoniae* protein of the  
 CC invention. The proteins (or their homologues, derivatives and/or  
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic  
 CC compositions comprising the proteins are useful as vaccines and also in  
 CC diagnostic assays. The sequences are useful for the detection or  
 CC diagnosis of *S. pneumoniae* infection, by contacting a sample to be tested  
 CC with them. Agents capable of antagonising, inhibiting or interfering with  
 CC the function or expression of the protein or polypeptide are useful in  
 CC medical compositions in the treatment or prophylaxis of *S. pneumoniae*  
 CC infection. As the sequences can be used to treat *S. pneumoniae* infection,  
 CC they can be used to treat bacterial pneumonia, which has high rates in  
 CC young children, the elderly, and in patients with predisposing conditions  
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,  
 CC or with immunosuppressive disorders, especially AIDS. They can also be  
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and  
 CC meningitis.

XX Sequence 2120 AA;

Query Match 100.0%; Score 4398; DB 21; Length 2120;  
 Best Local Similarity 100.0%; Pred. No. 6.4e-59;  
 Matches 664; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YPVVLADTSSSEDALNLSDEKVAENKEKHENHSAMETSODFKEKTAIVIKEKEVSKN 60  
 DB 1 YPVVLADTSSSEDALNLSDEKVAENKEKHENHSAMETSODFKEKTAIVIKEKEVSKN 60  
 QY 61 PVIDNNTSNEBAKIKEENSNKSGDYTDSFVNKNTENPKKEDKVYVIAEFKDSGEKAI 120  
 DB 61 PVIDNNTSNEBAKIKEENSNKSGDYTDSFVNKNTENPKKEDKVYVIAEFKDSGEKAI 120  
 QY 121 KELSLKNTKVLTYTDRIFNLSAIEETPDNDLKIQIIGISSVERAQVQPMNHARKKI 180  
 DB 121 KELSLKNTKVLTYTDRIFNLSAIEETPDNDLKIQIIGISSVERAQVQPMNHARKKI 180

QY 181 GVEAIDYLSKINAPFGKNPGRGMVTSNIDTGDYRHKAMRIDDDAKASRFKKEDLKG 240  
 DB 181 GVEAIDYLSKINAPFGKNPGRGMVTSNIDTGDYRHKAMRIDDDAKASRFKKEDLKG 240  
 QY 241 TDKRYMISDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNDEODIKNPNK 300  
 DB 241 TDKRYMISDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNDEODIKNPNK 300  
 QY 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGLVGEKY 360  
 DB 301 IDGIAPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGLVGEKY 360  
 QY 361 QAIRALRKAGIPMVVATGNVATSSASSSMDLVANNHLMKMTDGNVTRTAHEDAIIVAS 420  
 DB 361 QAIRALRKAGIPMVVATGNVATSSASSSMDLVANNHLMKMTDGNVTRTAHEDAIIVAS 420  
 QY 421 AKNOTVEFDKVNIGESFKRNIGAFDPSKITTNEDGTAPSLKKEVYIGKGDODLIG 480  
 DB 421 AKNOTVEFDKVNIGESFKRNIGAFDPSKITTNEDGTAPSLKKEVYIGKGDODLIG 480  
 QY 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGAIAIMVNTVYNNHNTTELPRMGYEADG 540  
 DB 481 LDLRGKIAVMDRIYTKDLKNAFKKAMDGAIAIMVNTVYNNHNTTELPRMGYEADG 540  
 QY 541 TKSQVPSISGDDGYKLMNMPDKTEVKRNKEDFKDKLEQYYPIMESFNSKPNVGD 600  
 DB 541 TKSQVPSISGDDGYKLMNMPDKTEVKRNKEDFKDKLEQYYPIMESFNSKPNVGD 600  
 QY 601 EKEIDFPAPPTDKELYKEDIIVPAGSTSWGPRIDLKLPVSAFGKIKSTLAVINGKS 660  
 DB 601 EKEIDFPAPPTDKELYKEDIIVPAGSTSWGPRIDLKLPVSAFGKIKSTLAVINGKS 660  
 QY 661 TYGY 664  
 DB 661 TYGY 664

RESULT 3  
 ID ABU01020 standard; Protein: 2140 AA.

XX ABU01020;  
 DT 11-FEB-2003 (first entry)  
 XX  
 DE S. pneumoniae type 4 strain protein from coding region #590.  
 XX  
 KM Bacterial meningitis; pneumonia; sepsis; otitis media;  
 KM ear infection; antiinflammatory; antibacterial; immunostimulant;  
 KM auditory; respiratory; gene therapy; vaccine.  
 XX  
 OS Streptococcus pneumoniae type 4 strain.  
 XX  
 PN W0200277021-A2.  
 PD 03-OCT-2002.  
 PF 27-MAR-2002; 2002MO-IB02163.  
 PR 27-MAR-2001; 2001GB-0007658.  
 PA (CHIR-) CHIRON SPA.  
 XX (GENO-) INST GENOMIC RES.  
 PI Maignant V, Tectelin H, Fraser C;  
 XX  
 DR WPI; 2003-040579/03.  
 DR N-PSDB; ABX06302.  
 XX  
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,  
 PT useful as medicaments for treating or preventing a disease or infection  
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media  
 PT or ear infection -

XX  
 PS Claim 1; SEQ ID No 1180; 56pp; English.  
 XX  
 CC The invention relates to a protein comprising or having at least 50%  
 CC identity to any of the 2469 amino acid sequences, identified in the  
 CC specification (available on a computer readable format), or its fragment,  
 CC expressed from 2469 of 2489 identified DNA coding regions from the  
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as  
 CC A8556454. Also included are an antibody which binds one of the  
 CC proteins, treating a patient by administering the protein, DNA or  
 CC antibody (in a composition), a kit comprising first and second primers,  
 CC which are the nucleic acid cited above or fragments between nucleotides  
 CC 8-100 of a sequence not defined in the specification, for amplifying a  
 CC target sequence contained within a Streptococcus nucleic acid sequence,  
 CC where the first primer is substantially complementary to the target  
 CC sequence and the second primer is substantially complementary to the  
 CC complement of the target sequence, and where the parts of the primers  
 CC having substantial complementarity define the termini of the target  
 CC sequence to be amplified, assay comprising contacting a test compound  
 CC with the protein, and determining whether the test compound binds to the  
 CC protein and a Streptococcus pneumoniae bacterium, where one or more  
 CC genes encoding the proteins has been rendered inactive. The proteins,  
 CC nucleic acid molecules, antibody and compositions are useful as  
 CC medicaments for treating or preventing a disease or infection due to  
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,  
 CC sepsis, otitis media or ear infection. They are also useful in developing  
 CC vaccines, diagnostics and antibiotics. The methods are useful for  
 CC identifying immunodominant proteins. The present sequence is one of  
 CC the 2469 proteins expressed by the identified coding regions from the  
 CC genomic sequence.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pcr\_sequences.  
 CC  
 XX  
 SQ Sequence 2140 AA;

Query Match 99.6%; Score 4379; DB 24; Length 2140;  
 Best Local Similarity 99.8%; Pred. No. 1, 3e-58;  
 Matches 662; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PVVLADTSSSEDALNISDKKVAENKKEHNIHSAMETSQDFPKKTAIVIKERVSNKP 61  
 DB 23 EVVLADTSSSEDALNISDKKVAENKKEHNIHSAMETSQDFPKKTAIVIKERVSNKP 62  
 QY 62 VVNNNTSNEBAKIKERNSKNSQGDYTDSPFNKNTENPKKEDKVVYIAEPDKEGSEKAIK 121  
 DB 83 VVNNNTSNEBAKIKERNSKNSQGDYTDSPFNKNTENPKKEDKVVYIAEPDKEGSEKAIK 142  
 QY 122 ELSSLKNTKVLKYDRIFNGSAIETTPDNDKIKOIEGSSVERAOKVOPMMHARKSIG 181  
 DB 143 ELSSLKNTKVLKYDRIFNGSAIETTPDNDKIKOIEGSSVERAOKVOPMMHARKSIG 202  
 QY 182 VBEAIDYLSKINAPFGKNPGRGMVTSNIDTGDYRHKAMRIDDDAKASRFKKEDLKG 241  
 DB 203 VBEAIDYLSKINAPFGKNPGRGMVTSNIDTGDYRHKAMRIDDDAKASRFKKEDLKG 262  
 QY 242 DKNYMLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNDEODIKNPNK 301  
 DB 263 DKNYMLSDKIPHAFFNYNGSKITVEKYDDGRDYFDPHGMHIAIGLGNDEODIKNPNK 322  
 QY 302 DGIAPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGLVGEKY 361  
 DB 323 DGIAPNAQIFSYKMYSDAGSGFAGDETFMFAIEDSIKHNVDVSVSSGFTGLVGEKY 382  
 QY 362 QAIRALRKAGIPMVVATGNVATSSASSSMDLVANNHLMKMTDGNVTRTAHEDAIIVAS 421  
 DB 383 QAIRALRKAGIPMVVATGNVATSSASSSMDLVANNHLMKMTDGNVTRTAHEDAIIVAS 442  
 QY 442 KNOTVEFDKVNIGESFKRNIGAFDPSKITTNEDGTAPSLKKEVYIGKGDODLIG 481  
 DB 443 KNOTVEFDKVNIGESFKRNIGAFDPSKITTNEDGTAPSLKKEVYIGKGDODLIG 502

```

Qy 482 DLRGKIAVMDBIYTKDLKNAFKKAMDKGARIMVNTVYNNRNMTLPAWGFEADSGT 541
Db 503 DLRGKIAVMDBIYTKDLKNAFKKAMDKGARIMVNTVYNNRNMTLPAWGFEADSGT 562
Qy 542 KSOVFSISGDDGVKLMNNINPDKTEVGRNKKEDFKDLGQYYPIDMSFNSKNVDE 601
Db 563 KSOVFSISGDDGVKLMNNINPDKTEVGRNKKEDFKDLGQYYPIDMSFNSKNVDE 622
Qy 602 KEIDFKFAPDIDKELYKEDIIVPAGSTSMGPRIIDLKPDVSAFGKNIKSTLVINGKST 661
Db 623 KEIDFKFAPDIDKELYKEDIIVPAGSTSMGPRIIDLKPDVSAFGKNIKSTLVINGKST 682
Qy 662 YGY 664
Db 683 YGY 685

```

RESULT 4  
AAW61246 standard; Protein; 1007 AA.

AAW61246;  
02-OCT-1998 (first entry)

Streptococcus pneumoniae SPI22 protein.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;  
detection; pneumonia; otitis media; meningitis.

Streptococcus pneumoniae.

MO9818930-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19422.

31-OCT-1996; 96US-0029960.

(HUMA-) HUMAN GENOME SCT INC.

Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

WPI, 1998-272224/24.

N-PSDB; AAV27431.

Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis

Claim 11; Page 92-93; 118pp; English.

The present sequence represents a protein from Streptococcus pneumoniae. The nucleic acid sequence encoding the Streptococcus pneumoniae protein can be useful in vaccines for inducing protective antibodies against Streptococcus pneumoniae, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect Streptococcus infection (by usual hybridisation or amplification methods), also for isolating Streptococcus genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or monitoring infections. Antibodies which bind the protein are used to detect corresponding antigens, to purify the protein and for passive immunisation (optionally coupled to a toxin). Vaccines are administered, e.g. by injection, orally or through the skin, typically at 0.01-1000 (especially 10-300) mu g/ml per dose.

Sequence 1007 AA;

Query Match 94.8%; Score 4171; DB 19; Length 1007;  
Best Local Similarity 100.0%; Pred. No. 3.8e-56;  
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 38 ETSODFEKKTAVIKEKEVSKNPVIDNNTSNEBAKIKEENSNSKOGDYTDSFVNKNTEN 97
Db 1 ETSODFEKKTAVIKEKEVSKNPVIDNNTSNEBAKIKEENSNSKOGDYTDSFVNKNTEN 60
Qy 98 PKKEDKVVYIAEFKOKSGEKALTELSSLKNTKYLYYDRIFNGSALETTPDNDLKKOI 157
Db 61 PKKEDKVVYIAEFKOKSGEKALTELSSLKNTKYLYYDRIFNGSALETTPDNDLKKOI 120
Qy 158 EGISVVERAOKVQPMNHARKEIGVEEAIDVLKSIAPFGNPGDGRGVNSIDTGDYR 217
Db 121 EGISVVERAOKVQPMNHARKEIGVEEAIDVLKSIAPFGNPGDGRGVNSIDTGDYR 180
Qy 218 HKAMRIDDDAKASRFKKEDLKGTDKNYWLSDKIPHAFNYNGSKITVEKYDDGRDYDP 277
Db 181 HKAMRIDDDAKASRFKKEDLKGTDKNYWLSDKIPHAFNYNGSKITVEKYDDGRDYDP 240
Qy 278 HGWHIAGILAGNDTEODIKNPNIGIDGIAIPNAQITSYKYSIDAGSGFAGDETFHAIEDSI 337
Db 241 HGWHIAGILAGNDTEODIKNPNIGIDGIAIPNAQITSYKYSIDAGSGFAGDETFHAIEDSI 300
Qy 338 KHNVDVSVSSGFTGTGLVGEKYNQAIIRALKAGIPMVVATGNATYSASSSSMDLVANNH 397
Db 301 KHNVDVSVSSGFTGTGLVGEKYNQAIIRALKAGIPMVVATGNATYSASSSSMDLVANNH 360
Qy 398 LKMTDTGNVTSTAHEDAIAVASAKNOTVEFDKYNIGESPKYNNIGAFPDKSKITTNE 457
Db 361 LKMTDTGNVTSTAHEDAIAVASAKNOTVEFDKYNIGESPKYNNIGAFPDKSKITTNE 420
Qy 458 GTKAPSKLKFVYIGKGDODLIGDLRGKIAVMDBIYTKDLKNAFKKAMDGARAIVVN 517
Db 421 GTKAPSKLKFVYIGKGDODLIGDLRGKIAVMDBIYTKDLKNAFKKAMDGARAIVVN 480
Qy 518 TVNYNNDNMTELPAWGFEADSGTKSOVFSISGDDGVKLMNNINPDKTEVGRNKKEDFK 577
Db 481 TVNYNNDNMTELPAWGFEADSGTKSOVFSISGDDGVKLMNNINPDKTEVGRNKKEDFK 540
Qy 578 DKLEQYYPIDMSFNSKNPNVGEDEKIDFKFAPDIDKELYKEDIIVPAGSTSMGPRIIDL 637
Db 541 DKLEQYYPIDMSFNSKNPNVGEDEKIDFKFAPDIDKELYKEDIIVPAGSTSMGPRIIDL 600
Qy 638 LKPDVSAFGKNIKSTLVINGKSTYGY 664
Db 601 LKPDVSAFGKNIKSTLVINGKSTYGY 627

```

RESULT 5  
ABP54664  
ID ABP54664 standard; Protein; 1007 AA.

ABP54664;

04-SEP-2002 (first entry)

S. pneumoniae SPI22 protein sequence SEQ ID NO:216.

Streptococcus pneumoniae; epitope; vaccine; antigenic protein;  
antibacterial; Streptococcal infection; detection.

Streptococcus pneumoniae.

US2002061545-A1.

23-MAY-2002.

22-JAN-2001; 2001US-0765272.

30-OCT-1997; 97US-0961083.

(CHOI/) CHOI G H.  
(KUNS/) KUNSCH C A.  
(BARA/) BARASH S C.  
(DILL/) DILLON P J.

PA (DOUG/) DOUGHERTY B.  
 PA (FRAN/) FANNON M R.  
 PA (ROSE/) ROSEN C A.  
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,  
 PI Rosen CA;  
 XX WPI; 2002-4/9261/51.  
 DR N-PSDB; ABQ84899.  
 XX  
 PT New Streptococcus pneumoniae antigens, useful for detecting  
 PT Streptococcus and for preventing or attenuating disease caused by  
 PT Streptococcus infection -  
 XX  
 PS Claim 11; Page 50; 70pp; English.  
 XX  
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the  
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.  
 CC The S. pneumoniae antigens have antibacterial activity and can be  
 CC used in vaccines. The S. pneumoniae antigens can also be used to  
 CC prevent or attenuate a Streptococcal infection in an animal. The  
 CC polynucleotides encoding the S. pneumoniae antigens can be used to  
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent  
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)  
 CC which are used in an example from the present invention.  
 CC  
 XX  
 SQ Sequence 1007 AA;  
 Query Match 94.8%; Score 4171; DB 23; Length 1007;  
 Best Local Similarity 100.0%; Pred. No. 3.8e-56;  
 Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 38 ETSQDFEKKTAIVKEKEVNSKNPVIINNTSNEBAKIKEENSKNSQGDYTDSPFNKNTEN 97  
 DB 1 ETSQDFEKKTAIVKEKEVNSKNPVIINNTSNEBAKIKEENSKNSQGDYTDSPFNKNTEN 60  
 QY 98 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNCSALETTPDNLDKIKOI 157  
 DB 61 PKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLTYDRIFNCSALETTPDNLDKIKOI 120  
 QY 158 EGISSVVERAKVQPMNMHARKEIGVEBAIDYLSINAPFGKPNPGRGMYISNIDTGYDR 217  
 DB 121 EGISSVVERAKVQPMNMHARKEIGVEBAIDYLSINAPFGKPNPGRGMYISNIDTGYDR 180  
 QY 218 HKAMRIDDDAKASMRFKKEDLKGTDKNYMLSDKIPHAFNYYNGGKIIVEXKDDGRDYFDP 277  
 DB 181 HKAMRIDDDAKASMRFKKEDLKGTDKNYMLSDKIPHAFNYYNGGKIIVEXKDDGRDYFDP 240  
 QY 278 HGHMIAIGILAGNDTEQDINKFNIGIDGIAPNAQIFSYMYSDAGSGFAGDETFPHAIEDSI 337  
 DB 241 HGHMIAIGILAGNDTEQDINKFNIGIDGIAPNAQIFSYMYSDAGSGFAGDETFPHAIEDSI 300  
 QY 338 KHAVNVVSVSGFTGTGLNCKEYQOALRALRKAGIPNVVATGNVATGATSSSSMDLVANNH 397  
 DB 301 KHAVNVVSVSGFTGTGLNCKEYQOALRALRKAGIPNVVATGNVATGATSSSSMDLVANNH 360  
 QY 398 LKMTDGTNVTGRTAHEDAIIVASAKNOTVEFDKNYIGESPKYKNISAFPDKSKITTNEED 457  
 DB 361 LKMTDGTNVTGRTAHEDAIIVASAKNOTVEFDKNYIGESPKYKNISAFPDKSKITTNEED 420  
 QY 458 GTKAPSKLTKFVYIGKGGDODLIGDLRGKIAVMDRITTKDLKNAFKKAMDKGARAIWVN 517  
 DB 421 GTKAPSKLTKFVYIGKGGDODLIGDLRGKIAVMDRITTKDLKNAFKKAMDKGARAIWVN 480  
 QY 518 TUNVYVNDNNTELPAWGYEADDEGTSQVFSISGDDGYKLMNMINPDCKTEYKRNKEDFK 577  
 DB 481 TUNVYVNDNNTELPAWGYEADDEGTSQVFSISGDDGYKLMNMINPDCKTEYKRNKEDFK 540  
 QY 578 DKLQEQYPIIMESFNKPNVYDEKEIDFKFAPDTEKELYKEDIIVPAGSTSMGPRIDL 637  
 DB 541 DKLQEQYPIIMESFNKPNVYDEKEIDFKFAPDTEKELYKEDIIVPAGSTSMGPRIDL 600  
 QY 638 LKPDVSAFGKNIKSTLVINGKSTYGY 664

DB 601 LKPDVSAFGKNIKSTLVINGKSTYGY 627.  
 RESULT 6  
 ID AAY86159 standard; Protein; 192 AA.  
 AC AAY86159;  
 DT 10-APR-2000 (first entry)  
 DE S. pneumoniae derived protein #368.  
 KM Treatment; prevention; disease; diagnosis; gene therapy; screening;  
 KM bacterial; antimicrobial; antibiotic; pathogenesis; infection.  
 OS Streptococcus pneumoniae.  
 PN WC9806734-A1.  
 PD 19-FEB-1998.  
 PF 15-AUG-1997; 97WO-US14436.  
 PR 16-AUG-1996; 96US-0024022.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,  
 PI Scodola RK;  
 DR WPI; 1998-159452/14.  
 DR N-PSDB; AA296473.  
 PT Streptococcus pneumoniae proteins and related DNA - useful for  
 PT screening compounds for antibacterial activity  
 PS Claim 5; Page 617; 640pp; English.  
 CC This invention describes novel isolated Streptococcus pneumoniae  
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see  
 CC AA296173-296494). The DNA, vectors and host cells described in the  
 CC method of the invention are useful for the recombinant expression of the  
 CC polypeptides. The polypeptides are useful for treatment or prevention of  
 CC disease, or diagnosis of disease related to expression or activity of  
 CC such a polypeptide. They can also be used to screen for compounds which  
 CC interact with and inhibit or activate such a polypeptide. The  
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful  
 CC for inducing an immunological response in a mammal. The antagonists are  
 CC useful to inhibit such bacterial polypeptides. The polypeptides are  
 CC particularly useful to identify antimicrobial compounds and antibiotics.  
 CC They are also useful to determine their role in pathogenesis of  
 CC infection, dysfunction and disease.  
 CC  
 XX  
 SQ Sequence 192 AA;  
 Query Match 22.1%; Score 974; DB 19; Length 192;  
 Best Local Similarity 94.7%; Pred. No. 1.2e-07;  
 Matches 161; Conservative 2; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 PVVLADTSSSEDLANTSDKEKVAENKKEHNIISAMETSODPKKTAIVKEKEVNSKNP 61  
 DB 23 EVLADTSSSEDLANTSDKEKVAENKKEHNIISAMETSODPKKTAIVKEKEVNSKNP 82  
 QY 62 VIDNNTSNEBAKIKEENSKNSQGDYTDSPFNKNTENPKKEDKVVYIAEFKDKESGEKAIK 121  
 DB 83 VIDNNTSNEBAKIKEENSKNSQGDYTDSPFNKNTENPKKEDKVVYIAEFKDKESGEKAIK 142  
 QY 122 ELGSLKNTKVLTYDRIFNCSALETTPDNLDKIKOEGISSVERAKVQ 171  
 DB 143 QLSLKNKTVLTYDRIFNCSALETTPDNLDKIKOEGISSVERAKVQ 192

```
RESULT 7
AAR10559
ID AAR10559 standard; Protein; 1962 AA.
XX
XX AAR10559;
AC
XX
XX 15-APR-1991 (first entry)
DT
XX Mutant protease (A137G/K138L/T139A).
DE
XX
XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
XX Lactococcus lactis SK11.
OS
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
XX Epa11715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI WPI; 1991-038622/06.
XX
XX N-PSDB; AAQ10413.
XX
XX Mutant protease gene(s) and protease(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
XX Diaclojure; Fig 1(1-7)+5(b); 29 pp; English.
XX
XX The mutant protease having new cleavage specificities is
XX obtained by replacing three amino acids.
XX This mutant may then be used to prepare hybrid proteases,
XX the fusion being between a type I and a type III protease of
XX L.lactis Wg2 and SK11.
XX The product has modified properties, e.g. thermostability,
XX alkaline/acid pH stability, oxidative stability, autoprolysis etc.,
XX compared to the parent protease(s). The proteases can be used for
XX preparing products (butter cheese, human and animal foodstuffs)
XX prepared with the aid of lactic acid bacteria.
XX See also AAQ10411-17 and AAQ10870-71.
XX
XX Sequence 1962 AA;
SO
Query Match 20.5%; Score 899.5; DB 12; Length 1962;
Best Local Similarity 25.9%; Pred. No. 0.00057;
Matches 180; Conservative 132; Mismatches 247; Indels 135; Gaps 30;
OY 1 YPV--VLADTSSSEDALNISDEKVAENKEKHENIHSAMETSODPEKKTVIKEKEVVS 58
DB 24 LPVGEIOAKKAIISQOTLSSSLANTYTA-----TKQAATDTTAAT----- 65
OY 59 KNPVIDNNTSNEAKIKENSNGQDYDTSFVNKNTENPKKEDYVYIAEFKESGEK 118
DB 66 -NQAIATGLAKGIDYNTKLVQODIYDVIVQM-SAAPASENGLTLTDVSSTAIEQOE 123
OY 119 AIKELISSIAKTKVLY-----TYRINFGSAIETTPNLDKIKIIEGISSVERAQK 168
DB 124 TNKVIAAGASVYAAVEQVTOGTAGESYGVVNGFSTKRVVVDIPRLKOIAGVKTTLAKV 183
```

```
OY 169 VQPMNHARKETIVEEALDYLSINAPFGKFDGRGNVISNIDTGTDRHKAIRIDDAK 228
DB 184 YPTDADAKANSMANVQAWSNK-----YKGEITVVSVIDSGIDPTTHDKMLSDXKD 234
OY 229 ASMRFKKEDL-KGTDK--NYWLSDKIPLAFNY-YNGSKITVEKYDGRDYFDHGMHIA 283
DB 235 V--KLTKSDVEKFPDVTWGRYFNSKVPYGFNVADNNDIT--DDKVD--EKGHMHVA 286
OY 284 GILAGNDTEODIKNFNGIDGIAFPAQIFSYVMYSDAG-SGPAGDETFPHAEISIKHVD 342
DB 287 GIGANGIGDDPA--KSVGVAPBAQLAMKVFNSNDSGTSLAGSATVVSALSDSAKGAD 344
OY 343 VVSVSG-FTGTGLVGEKYMQAIRALKRAGIPMVVATGNVATSSASSMDLVANNHLKMT 401
DB 345 VLNMISLGSNSGNQLEDPBELAAVGNANESGPAVIAISGNSGTSGSATG--VNDYVGLQ 402
OY 402 DTGNVTRTAHEDALVANSKNOVEFDKNI-----GGSEFK--RRTGAFPEKSK 451
DB 403 DNEWVGSPGTSRGATVVASAENTVITQAVTITDGTGLQLPPTIOLSHDFTSGFDQK 462
OY 452 ITTNEBDGKAPSKLFVYIGKGODDILGLDLRGKIAVMDRI-YTKDLKNAFKKAMDGA 510
DB 463 FYIVKDSGN-----LSKGLADYTA--DAKGIKIVKRGERSFDDKQKTAQDA--GA 511
OY 511 RAINVNTVNYNNDNMTELPAMGYEADBGTSQVFSISGDDGYKL--WNMINPDKTEV 568
DB 512 AGLIIVNT-----DGTATPMT--STALTTPTFGSSVTGKLVDMVTAHPDDSLGV 562
OY 569 KRNNKEDFKDLE-QYPIPIMESFNSKPNVGDKEKIDFKFAPDTXELYKEDIIVPAGS 627
DB 563 KIT-----LAMLPMQKYTEBKMS-----DF----- 582
OY 628 TSWGPRIDLKPPVSAFGKNIKSTLVNKGST 661
DB 583 TSYGVSNLSFKPDITAPGGINWSTON--NNGYTN 615
RESULT 8
AAR10557
ID AAR10557 standard; Protein; 1962 AA.
XX
XX AAR10557;
AC
XX
XX 15-APR-1991 (first entry)
DT
XX
XX Mutant protease (A137G/K138D).
XX
XX Mutant protease gene; fermentation; foodstuff; flavouring;
KW lactic acid bacteria.
XX
XX Lactococcus lactis SK11.
OS
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
XX Epa11715-A.
XX
XX 06-FEB-1991.
XX
XX 02-AUG-1990; 90EP-0202113.
XX
XX 04-AUG-1989; 89NL-0002010.
XX
XX (NEZU-) NED INST ZUIVELONDE.
XX
XX Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;
PI WPI; 1991-038622/06.
XX
XX N-PSDB; AAQ10411.
```





```

Db      124  TNKVIAQASVKAAVEQVTQGTAGESYGVVNGSTKVRVVDIPKQLQIAGVTVTLAKV 183
Qy      169  VQPMNHNARKEIGVEEADIDYLSINAPFGKNFDRGWNISNIDGTDRYRHKAMRIDDDAK 228
      184  YPPTDAKANSMAVQAQVMSNTK-----YKSGEIVASVIDSGIDPETHDKMLSDDXD 234
Qy      229  ASMRFFKEDL-KGTDK--NYWLSDKIPIHAFNY-VNGSKITVEKYDGRDYFDPHGMIIA 283
      235  V--KLTKSDVEKFTDVTGHRVYFNSKVYGFYVADNDNTIT---DDKVD--EQGHMIVA 286
Qy      284  GILAGNDEODIKNPNNGIDGIAAPNAQIFSYKMYSD-----AG-SGFAGDETFHAIEDS 336
      287  GIIANGTGDDPA--KSVVGVAPEAQILAMKVFNSDTSAGATTGTGSAIVVSAIEDS 344
Qy      337  IKNHVDVVS-VSG-FTGTGLVGEKYOQAIRALKRAGIPMVVATGNYATSASSSMDLVAN 395
      345  AKIGADVAMSLGNSNGOTLEDEPRLAIVONANESGTAAVISAAGNSGTSGSATEB--VNK 402
Qy      396  NHLKMTDTGNVTRTAHEDAIIVASAKNQVEFDKVI-----GGESFKY--RNIGA 445
      403  DYVGLODNEWVGSPGTSRGATTVASAENTDVIYQAVITITDGTGLQIGPETIQLSHDPTG 462
Qy      446  FPDKSKITTNDDGTAPSKLKFVYIGKQDODIGLDKRTIAVMDRI-YTKDLKNAKK 504
      463  SFDQKKFYIVKASGN-----LSKALADYTA-DAKGKIALVKRGEFFDDKQKXAAQ 513
Qy      505  AMDKGAIRIMVNTVYVNRDNMTLPAWGEADEGTSQVFSISGDDGVKL--WNMINP 562
      514  AA--GAALIIVNT-----DGTATPMT--SIALTTTPFTGLSSVTGQLVDMVTAHP 562
Qy      563  DKTEVKNKKEDFKDLE-QYYPIDMESFNSNKNPNVGEDEKIDFKFAPDTKELYKEDI 621
      563  DDSLGVKIT-----LAMLPNQKYTEDKMS-----DF----- 588
Qy      622  IVPAGSTWGPRIIDLKPDVASAPGNKIKSTLVNINKST 661
      589  -----TSYGPVSNLSFKPDITAPGGINWSTON--NNGYTN 621

```

## RESULT 10

AAR10558 standard; Protein; 1962 AA.

AAR10558;

15-APR-1991 (first entry)

Mutant protease (A137G/K138P/T139P).

Mutant protease gene; fermentation; foodstuff; flavouring;

lactic acid bacteria.

Lactococcus lactis SK11.

Key Location/Qualifiers

Peptide 1..187

Protein /label= signal\_peptide

188..1962

Protein /label= mature\_protein

BP111715-A.

06-FEB-1991.

02-AUG-1990; 90BP-0202113.

04-AUG-1989; 89NL-0002010.

(NEZU-) NED INST ZUIVELONDE.

Vos PAJ, Siezen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ;

WPI, 1991-038622/06.

```

DR      N-PSDB; AAQ10412.
XX      Mutant protease gene(s) and protease(s) - derived from type I and
PT      III protease genes from lactococcal strains, used in fermentation
PT      foodstuffs and flavourings
XX
XX      Disclousure; Fig 1(1-7)+5(b); 29 pp; English.
PS
PS      The mutant protease having new cleavage specificities is
CC      obtained by replacing three amino acids.
CC      This mutant may then be used to prepare hybrid proteases,
CC      the fusion being between a type I and a type III protease of
CC      L.lactis Wg2 and SK11.
CC      The product has modified properties, e.g. thermostability,
CC      alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC      compared to the parent protease(s). The proteases can be used for
CC      preparing products (butter cheese, human and animal foodstuffs)
CC      prepared with the aid of lactic acid bacteria.
CC      See also AAQ10411-17 and AAQ10870-71.
CC
SQ      Sequence 1962 AA;

```

Query Match 20.2%; Score 888.5; DB 12; Length 1962;  
 Best Local Similarity 25.8%; Pred. No. 0.00084;  
 Matches 179; Conservative 131; Mismatches 249; Indels 135; Gaps 30;

```

Qy      1  YPV--VLADTSSSEDAINISDKEKVAENKEHNIHAMETSODFKKTAIVIKEVVS 58
      24  LPVEIQAKAIAISQTLGSSLANVTAA-----TAKQAATTTTAAIT----- 65
Db
Qy      59  KNPIVDNNTSNEAKIKENSNSQGYTDSFVNKNTENPKEDKVVYVIAEFKKEGSEK 118
      66  -NQAIAVQLAAKIDYNNKLVQOQDIYVDYIVQM--SAAPASENGIIRTDYSSTAIEIOE 123
Db
Qy      119  AIKELSLKNTKLVY-----TYDRIFNCSAIIPTPNLMDKIOEGISSVERAQ 168
      124  TNKVIAQASVKAAVEQVTQGTAGESYGVVNGSTKVRVVDIPKQLQIAGVTVTLAKV 183
Qy      169  VQPMNHNARKEIGVEEADIDYLSINAPFGKNFDRGWNISNIDGTDRYRHKAMRIDDDAK 228
      184  YPPTDAKANSMAVQAQVMSNTK-----YKSGEIVASVIDSGIDPETHDKMLSDDXD 234
Db
Qy      229  ASMRFFKEDL-KGTDK--NYWLSDKIPIHAFNY-VNGSKITVEKYDGRDYFDPHGMIIA 283
      235  V--KLTKSDVEKFTDVTGHRVYFNSKVYGFYVADNDNTIT---DDKVD--EQGHMIVA 286
Db
Qy      284  GILAGNDEODIKNPNNGIDGIAAPNAQIFSYKMYSDAG--SGFAGDETFHAIEDSIKHNVD 342
      287  GIIANGTGDDPA--KSVVGVAPEAQILAMKVFNSDTSAGATTGTGSAIVVSAIEDSAKIGAD 344
Db
Qy      343  VVSVSSG-FTGTGLVGEKYOQAIRALKRAGIPMVVATGNYATSASSSMDLVANNHLMKT 401
      345  VLNMSLSGNSNGOTLEDEPRLAIVONANESGTAAVISAAGNSGTSGSATEG--VNDYVQLQ 402
Qy      402  DTGNVTRTAHEDAIIVASAKNQVEFDKVI-----GGESFKY--RNIGAEPDKSK 451
      403  DNEWVGSPGTSRGATTVASAENTDVIYQAVITITDGTGLQIGPETIQLSHDPTGSPDQKK 462
Db
Qy      452  ITTNEDEGKAPSKLKFVYIGKQDODIGLDKRTIAVMDRI-YTKDLKNAFKKAMDKGA 510
      463  FYIVKASGN-----LSKALADYTA-DAKGKIALVKRGEFFDDKQKQAQAA--GA 511
Db
Qy      511  RAIMVNTVYVNRDNMTLPAWGEADEGTSQVFSISGDDGVKL--WNMINPDKTEV 568
      512  AGLIIVNT-----DGTATPMT--SIALTTTPFTGLSSVTGQLVDMVTAHDDDSLGV 562
Qy      569  KRNKKEDFKDLE-QYYPIDMESFNSNKNPNVGEDEKIDFKFAPDTKELYKEDIIVPAGS 627
      563  KIT-----LAMLPNQKYTEDKMS-----DF----- 582
Db
Qy      628  TSGGPRIIDLKPDVASAPGNKIKSTLVNINKST 661
      583  TSYGPVSNLSFKPDITAPGGINWSTON--NNGYTN 615
Db

```



[illegible][illegible]

```

Db      124 TNKVIAAQASVKAABEVOVTQOTAGESYGVVNGFSTKRVVDIEPKLQIAGVKTVTLAKV 183
Qy      169 VQPMNHARKEIGVEBEAIDYLSINAPFGKNFDRGCVINISIDPTGYRHRKAMRIDDPAK 228
Db      184 YYPIDAKANSMAVNAQAVWSNKK-----YKGEQIVSVYIDSGIDPTKDMRLSDDKD 234
Qy      229 ASMRFKKEDL-KGTDK--NYWLSDKIPHAENV-YNGSKITVEKYDGRDYPDPHGMHIA 283
Db      235 V--KLTKSDVEKFTDVTWVGHRFYNSKVPYGFNVADNNDIT---DDKVD--EQHGMHVA 286
Qy      284 GILAGNTEQDIKNFNGIDGIAFPAQIFSYKMYSDAG--SGFAGDETFPHAIEDSIKXANVD 342
Db      287 GILGANGTGDDPA--KSVVGVAPEAQLAMKVFNSNSTSANTGSATVVSALIEDSAKIGAD 344
Qy      343 VVSVSSGF-TGTGLVGEKYMQAIRALRKAGIPMVVATGNVATSSSSMDLVANNHLKMT 401
Db      345 VLNMSLSGSDSGNQTLDEBELAAVONANESGTAIVISAGNSGTSAGTEG--VKKDYVGLQ 402
Qy      402 DTGNVTRTAHEDAIIVASAKNQTFEEDKVN-----GGESEPKY--RNIGAFPDPKSK 451
Db      403 DNEWVSGPSTSRGATVVAASAEENTDVIITQAVTITDGTGLQIGPETIQLSHDFTGSPQOK 462
Qy      452 ITTNEDGTAPSKLKFVYIGGQDODLIGDLRGKIAVMORI-YTKDLKNAFKKAMDKGA 510
Db      463 FYIVKQASGN-----LSKGLADYTA--DAKGIKAIIVKRGESFPDDKOKYQAQA--GA 511
Qy      511 RAIMVNVTVVYVNRDNWTELPAMGYBADEGTSQVFSISGDDGVKL--MMNINPDKKTREV 568
Db      512 AGLIIVNT-----DGTATPMT--SIALTTTFPTGLSSVTGQKLVDTWTAHPDSDSLGV 562
Qy      569 KRNKKEPDFOKLE-QYYPIDMESFNSNKPVNGEKEIDPFKAPDTELKELYKEDIIVPAGS 627
Db      563 KIT-----LAMLPNQKYTEDKMS-----DF-----582
Qy      628 TSWGPRIDLILKPDVSAFGKNIKSTLVINGKST 661
Db      583 TSYGPAVNSLSFKPDITAPGGINWSTON--NNGYTN 615

RESULT 14
AARI0563
ID AARI0563 standard; Protein; 1962 AA.
XX
AC AARI0563;
XX
DT 15-Apr-1991 (first entry)
XX
DE Mutant protease (K748T).
XX
KM Mutant protease gene; fermentation; foodstuff; flavouring;
XX
KM lactic acid bacteria.
XX
OS Lactococcus lactis SK11.
XX
FH Key Location/Qualifiers
FT Peptide 1..187
FT /label= signal_peptide
FT Protein 188..1962
FT /label= mature_protein
XX
PN BP411715-A.
PD 06-FEB-1991.
XX
PR 02-AUG-1990; 90EP-0202113.
XX
PR 04-AUG-1989; 89NL-0002010.
XX
PA (NEZU-) NED INST ZUIVELONDE.
XX
PI Vos PAJ, Slezzen RJ, De Vos WM, Kok J, Venema G, Haandrikman AJ,

```

```

DR WPI, 1991-038622/06.
DR N-PSDB; AAQ10417.
XX
PT Mutant protease gene(s) - derived from type I and
PT III protease genes from lactococcal strains, used in fermentation
PT foodstuffs and flavourings
XX
PS Disclosure; Fig 1(1-7)+5(a); 29 pp; English.
XX
CC The mutant protease K748T having new cleavage specificities is
CC obtained by carrying out single amino acid substitutions.
CC This mutant may then be used to prepare hybrid proteases,
CC the fusion being between a type I and a type III protease of
CC L. lactis Wg2 and SK11.
CC The product has modified properties, e.g. thermostability,
CC alkaline/acid pH stability, oxidative stability, autolysis etc.,
CC compared to the parent protease(s). The proteases can be used for
CC preparing products (butter cheese, human and animal foodstuffs)
CC prepared with the aid of lactic acid bacteria.
CC See also AAQ10411-17 and AAQ10870-71.
XX
SQ Sequence 1962 AA.

Query Match 20.1%; Score 885.5; DB 12; Length 1962;
Best Local Similarity 25.6%; Pred. No. 0.00094;
Matches 178; Conservative 132; Mismatches 249; Indels 135; Gaps 30;

Qy 1 YPV--VLADTSSSEDLNIDSKERVAENKEKHENHSAMETSODFEKETAIVKEKEVVS 58
Db 24 LPVGEIQAQKAIQOQTGLSSLVNTVTA-----TAKAATDTYTAATV-----65
Qy 59 KNPVIDNNTSNEBAKIKESNSKSGDYDTSFVNKATENPKCKDKVYVIAFPKDESGEK 118
Db 66 -NGAINTQLAKIDVYKLNKVKQOQDIYDVYIQW-SAAPASNGILRTYTSSTAIRIQE 123
Qy 119 AIRELSLKTKTVL-----TYDRIENGSAIETTPDNDLKIKQIEGISSVERAQK 168
Db 124 TNKVIAAQASVKAABEVOVTQOTAGESYGVVNGFSTKRVVDIEPKLQIAGVKTVTLAKV 183
Qy 169 VQPMNHARKEIGVEBEAIDYLSINAPFGKNFDRGCVINISIDPTGYRHRKAMRIDDPAK 228
Db 184 YYPIDAKANSMAVNAQAVWSNKK-----YKGEQIVSVYIDSGIDPTKDMRLSDDKD 234
Qy 229 ASMRFKKEDL-KGTDK--NYWLSDKIPHAENV-YNGSKITVEKYDGRDYPDPHGMHIA 283
Db 235 V--KLTKSDVEKFTDVTWVGHRFYNSKVPYGFNVADNNDIT---DDKVD--EQHGMHVA 286
Qy 284 GILAGNTEQDIKNFNGIDGIAFPAQIFSYKMYSDAG--SGFAGDETFPHAIEDSIKXANVD 342
Db 287 GILGANGTGDDPA--KSVVGVAPEAQLAMKVFNSNSTSANTGSATVVSALIEDSAKIGAD 344
Qy 343 VVSVSSGF-FGTGTLVGEKYMQAIRALRKAGIPMVVATGNVATSSSSMDLVANNHLKMT 401
Db 345 VLNMSLSGSDSGNQTLDEBELAAVONANESGTAIVISAGNSGTSAGTEG--VKKDYVGLQ 402
Qy 402 DTGNVTRTAHEDAIIVASAKNQTFEEDKVN-----GGESEPKY--RNIGAFPDPKSK 451
Db 403 DNEWVSGPSTSRGATVVAASAEENTDVIITQAVTITDGTGLQIGPETIQLSHDFTGSPQOK 462
Qy 452 ITTNEDGTAPSKLKFVYIGGQDODLIGDLRGKIAVMORI-YTKDLKNAFKKAMDKGA 510
Db 463 FYIVKQASGN-----LSKGLADYTA--DAKGIKAIIVKRGESFPDDKOKYQAQA--GA 511
Qy 511 RAIMVNVTVVYVNRDNWTELPAMGYBADEGTSQVFSISGDDGVKL--MMNINPDKKTREV 568
Db 512 AGLIIVNT-----DGTATPMT--SIALTTTFPTGLSSVTGQKLVDTWTAHPDSDSLGV 562
Qy 569 KRNKKEPDFOKLE-QYYPIDMESFNSNKPVNGEKEIDPFKAPDTELKELYKEDIIVPAGS 627
Db 563 KIT-----LAMLPNQKYTEDKMS-----DF-----582
Qy 628 TSWGPRIDLILKPDVSAFGKNIKSTLVINGKST 661

```

